

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 15:20:44 ; Search time 417.308 seconds
(without alignments)
3115.905 Million cell updates/sec

Title: US-10-676-299-3

Perfect score: 30

Sequence: 1 ttaatcatatgcgtttttgttgatgttg 30

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pi.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	30	100.0	727	1	ECRFARSR	X16045 E. coli R-f
2	28.4	94.7	120826	1	AF005147	AF005147 Salmonella
3	23.2	77.3	177109	2	AC118161	AC118161 Rattus no
4	23.2	77.3	159985	2	AC130151	AC130151 Rattus no
5	22	73.3	152772	9	AC099326	AC099326 Homo sapi
6	22	73.3	177672	2	AC010959	AC010959 Homo sapi
7	22	73.3	183748	9	AC124915	AC124915 Homo sapi
8	22	73.3	207693	2	AC108840	AC108840 Mus muscu
9	22	73.3	223489	5	EX248100	EX248100 Zebrafish
10	22	73.3	295094	2	EX649429	EX649429 Danio rer
11	21.6	72.0	6391	6	AX323523	AX323523 Sequence
12	21.6	72.0	10099	1	AE002489	AE002489 Neisseria
13	21.6	72.0	148151	9	AL513533	AL513533 Human DNA
14	21.6	72.0	169613	2	BX321886	BX321886 Danio rer
15	21.6	72.0	174303	10	AC144801	AC144801 Mus muscu
16	21.6	72.0	182051	9	AC007638	AC007638 Homo sapi
17	21.6	72.0	207265	2	AC122227	AC122227 Mus muscu
18	21.6	72.0	214815	2	AC130222	AC130222 Mus muscu
19	21.6	72.0	221911	2	AC114095	AC114095 Rattus no
20	21.6	72.0	241420	2	AC094501	AC094501 Rattus no
21	21.6	72.0	349980	6	AX044032	AX044032 Sequence
22	21.2	70.7	73000	2	AC090562	AC090562 Homo sapi
23	21.2	70.7	88747	9	AC104409	AC104409 Homo sapi
24	21.2	70.7	160811	9	AC020553	AC020553 Homo sapi
25	21.2	70.7	184897	9	AC090919	AC090919 Homo sapi
26	21	70.0	1465	5	BC049518	BC049518 Danio rer
27	21	70.0	2664	9	D89962	D89962 Homo sapien
28	21	70.0	4133	9	AB032158	AB032158 Homo sapi
29	21	70.0	65542	9	AC091766	AC091766 Homo sapi
30	21	70.0	110685	10	AL928957	AL928957 Mouse DNA
31	21	70.0	131289	2	AC147278	AC147278 Pan trogl
32	21	70.0	137156	2	AC138846	AC138846 Homo sapi
33	21	70.0	140410	2	AC139483	AC139483 Homo sapi
34	21	70.0	145692	9	AL355303	AL355303 Human DNA
35	21	70.0	148385	9	HSJ520B18	AL121978 Human DNA
36	21	70.0	150959	2	BX322574	BX322574 Danio rer
37	21	70.0	151259	2	AC138822	AC138822 Homo sapi
38	21	70.0	154803	9	AC131392	AC131392 Homo sapi
39	21	70.0	157792	9	AC069209	AC069209 Homo sapi
40	21	70.0	158755	2	AC145032	AC145032 Homo sapi
41	21	70.0	160127	2	AC068977	AC068977 Homo sapi
42	21	70.0	160701	9	AC108106	AC108106 Homo sapi
43	21	70.0	160901	2	AC138812	AC138812 Homo sapi
44	21	70.0	161000	2	AC145100	AC145100 Homo sapi
45	21	70.0	161362	2	AL355503	AL355503 Homo sapi

ALIGNMENTS

RESULT 1
ECRFARSR ECRFARSR 727 bp DNA linear BCT 07-SEP-1994
LOCUS E. coli R-factor R773 arsr gene.
DEFINITION X16045
ACCESSION X16045
VERSION X16045.1 GI:42716
KEYWORDS arsenical resistance; arsr gene; Arsr protein; DNA-binding protein;
regulatory protein; resistance gene.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 727)
AUTHORS San Francisco, M.J., Hope, C.L., Owolabi, J.B., Tisa, L.S. and
Rosen, B.P.


```

MEDLINE      98268996
PUBMED      9603870
REFERENCE   13
AUTHORS    Komano,T., Yoshida,T., Narahara,K. and Furuya,N.
TITLE      The transfer region of IncII plasmid R64: similarities between R64
           tra and legionella lcm/dot genes
JOURNAL    Mol Microbiol. 35 (6), 1348-1359 (2000)
MEDLINE    20223621
PUBMED     10760136
AUTHORS    Sampei,G., Komano,T., Sasaki,T., Tachibana,K., Furuya,N., Saito,Y.,
           Suzuki,T. and Mizobuchi,K.
TITLE      Organization and diversification of plasmid genomes: complete
           nucleotide sequence of R64 genome
JOURNAL    Unpublished
REFERENCE   15 (bases 1 to 120826)
AUTHORS    Komano,T., Sasaki,T., Tachibana,K., Furuya,N., Saito,Y., Suzuki,T.,
           Sampei,G. and Mizobuchi,K.
TITLE      Direct Submission
JOURNAL    Submitted (07-MAY-2002) Gen-ichi Sampei, The University of
           Electro-Communications, Applied Physics and Chemistry: Chofugaoka
           1-5-1, Chofu, Tokyo 182-8585, Japan (E-mail:sampeig@pc.uec.ac.jp,
           Tel:81-424-43-5481, Fax:81-424-43-5501)
FEATURES   Location/Qualifiers
           1..120826
           /organism="Salmonella typhimurium"
           /mol_type="genomic DNA"
           /db_xref="taxon:602"
           /lab_host="Escherichia coli strain K-12"
           /plasmid="R64"
           /note="R64 strain drd-11"
           complement(join(20047..20418,3609..4147))
           /note="100 pct identical to sp:Y122_ECOLI[hypothetical
           protein of insertion sequence IS2]
           location complement(4143..4147) and location
           complement(20047..20051) are duplicated"
           378..467
           /gene="repY"
           378..467
           /gene="repY"
           /note="100 pct identical to pit:A35445[RepY of plasmid
           Colib-P9]
           possible regulatory reading frame"
           /codon_start=1
           /transl_table=11
           /protein_id="BAB91567.1"
           /db_xref="GI:20521503"
           /translation="MKPQRFNPVQCINTRNSAISDSLWQV"
           455..1486
           /gene="repZ"
           455..1486
           /gene="repZ"
           /note="100 pct identical to sp:REPZ_ECOLI,sp:P18023[repZ
           of plasmid Colib-p9]"
           /codon_start=1
           /transl_table=11
           /protein_id="BAB91568.1"
           /db_xref="GI:20521504"
           /translation="MAGLKNTPYNAVHWSQLAPEQIRPWEDYRAGRATFLVPERK
           RKRRGHSHTKPCENSWYRPERKALQGLGHAYNLVKDQPTVTCGLRMMSR
           HPFYQKRTFVGRKAFPRKORLLDAIPVLVSPSDAGTHTVGMSVTRLABETSPKD
           SEGHWIPELVTSVSRLLAEQVPEGLVGYSEETMDREHQRIPRVVWITPAGQM
           LGVDWKLHQOQKRLRSEIRQQLIREGLREDEDSVHAARKWYLQRSQDALKR
           RKAASKRANKKKLPVQIQIYEMAEYLRKLPDPEAFYCSDDHLKRLAIRLRQLE
           LTLAAPP"
           2761..3114
           /gene="arsR"
           2761..3114
           /gene="arsR"
           /note="100 pct identical to sp:ARR1_ECOLI,sp:P15905[arsR
           of plasmid R773]"
           /codon_start=1
           /transl_table=11
           /protein_id="BAB91569.1"
           /db_xref="GI:20521505"
           /translation="MLQITFLQPKNISDETRILVILLRMGSLCVCDCMLDQSQ
           PKISRHLMRLRESGILLDRKQGWVHYRLSPHPSWAQIIIEQAWLSQDDVQVIARK
           LASVNCSSGSKAVCI"
           3162..3524
           /gene="arsD"
           3162..3524
           /gene="arsD"
           /note="99 pct identical to sp:ARD1_ECOLI,sp:P46003[arsD of
           plasmid R773]"
           /codon_start=1
           /transl_table=11
           /protein_id="BAB91570.1"
           /db_xref="GI:20521506"
           /translation="MKTLMFDPAMCSTGCGTDFVQALVDPADYQWLKCCGVOIE
           RFNLAQPMFQVNEKVKAFIASGAGLPLLLDGETVMAGRIYFKRABLARWFGIPL
           DKVLAPSGCGGNTSCC"
           join(3542..3597,20833..22533)
           /gene="arsA"
           join(3542..3597,20833..22533)
           /gene="arsA"
           /note="99 pct identical to pit:A25937[arsA of plasmid
           R773, arsenical pump-driving Appase]
           location 3593..3597 and location 20833..20837 are
           duplicated"
           join(3598..4147,20047..20832)
           /insertion_seq="IS2"
           complement(join(3609..4147,20047..20418))
           /gene="yadA"
           join(4148..7818,18208..20046)
           /note="different from Th5393[gb:M96392] in IS1133 insert
           position and terminal direct repeat"
           Query Match      94.7%; Score 28.4; DB 1; Length 120826;
           Best Local Similarity 96.7%; Pred. No. 0.74;
           Matches 29; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TTAATCATATGCGTTTGGTTATGTGTTG 30
       |||||
Db      2678 TTAATCATATGCGTTTGGTTATGTGTTG 2707

RESULT 3
AC118161/c
LOCUS      AC118161
DEFINITION Rattus norvegicus clone CH230-296C3, WORKING DRAFT SEQUENCE.
ACCESSION AC118161
VERSION   AC118161.6 GI:25099932
KEYWORDS  HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE    Rattus norvegicus
ORGANISM  Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE 1 (bases 1 to 177109)
AUTHORS  Murny,D., Maris., Metzker,M., Lee., Abramzon,S., Adams,C., Alder,J.,
           Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,B.,
           Anyalbechei,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
           Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
           Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
           Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
           Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,
           Chacks,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
           Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
           Davila,M.B., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
           Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
           Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
           Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
           Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
           Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
           Gebregeorgis,B., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
           Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,

```

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpaty, S., Kelly, S., Khan, Z., King, Z., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, M., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulsegged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakoelameh, O., Okwuon, G., Olarnpunaagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Poinexter, A., Popovic, D., Primus, E., Pu, L.-L., Puaro, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, F., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V., Yu, P., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 177109)
Worley, K.C.
Direct Submission
Submitted (14-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 177109)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:22856412.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GVAV
Center clone name: CH230-296C3
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 172459 bases at least Q40
Consensus quality: 173621 bases at least Q30
Consensus quality: 174370 bases at least Q20
Estimated insert size: 177946; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 177109: contig of 177109 bp in length.
FEATURES
source
Location/Qualifiers
1..177109
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-296C3"
misc_feature
1..1326
/note="wgs_contig"
ORIGIN
Query Match 77.3%; Score 23.2; DB 2; Length 177109;
Best Local Similarity 89.3%; Pred. No. 64;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 1 TTATCATATGCGTTTGGTATGCT 28
|||||
Db 79960 TTATCATATGCGTTTGGTATGCT 79933
|||||

AC130151 199985 bp DNA linear HTG 10-MAY-2003
AC130151 Rattus norvegicus clone CH230-42L16, *** SEQUENCING IN PROGRESS
***, 6 unordered pieces.
AC130151.4 GI:30522322
HTG: HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Rutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 199985)
Muzny, D., Marie, Metzker, M., Lee, A., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Ayalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabis, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpaty, S., Kelly, S., Khan, Z., King, Z., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, M., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulsegged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, S., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, P., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puro, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taboz, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Vadász, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 199985)
Worley, K.C.

Direct Submission
Submitted (08-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 199985)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:25009137. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GENC
Center clone name: CH230-42L16
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 130611 bases at least Q40
Consensus quality: 136437 bases at least Q30
Consensus quality: 139728 bases at least Q20
Estimated insert size: 141175; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 170190: contig of 170190 bp in length
* 170191 170290: gap of unknown length
* 170291 192968: contig of 22678 bp in length
* 192969 193068: gap of unknown length
* 193069 194488: contig of 1420 bp in length
* 194489 194588: gap of unknown length
* 194589 195684: contig of 1096 bp in length
* 195685 195784: gap of unknown length
* 195785 197169: contig of 1385 bp in length
* 197170 197269: gap of unknown length
* 197270 199985: contig of 2716 bp in length.

FEATURES
source
1..199985
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-42L16"
misc_feature
106671..108418
misc_feature
170291..172609
/note="wgs contig"
/note="wgs_contig"

ORIGIN

Query Match 77.3%; Score 23.2; DB 2; Length 199985;
Best Local Similarity 89.3%; Pred. No. 63;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTAATCATATCGTTTGTGATATGTGT 28

DB 180809 TTAATCATATCGTTTGTGATATGTGT 180836

RESULT 5

AC099326 152772 bp DNA linear PRI 24-JAN-2002
LOCUS Homo sapiens chromosome 3 clone RP11-7B12, complete sequence.
DEFINITION AC099326 AC010957
ACCESSION AC099326.1 GI:16874860
VERSION
KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152772)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.

Direct Submission

Unpublished

2 (bases 1 to 152772)

Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.

Direct Submission

Submitted (09-NOV-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA

3 (bases 1 to 152772)

Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.

Direct Submission

Submitted (24-JAN-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA

On Nov 9, 2001 this sequence version replaced gi:10198411.

COMMENT

----- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: <http://www.genome.washington.edu>
Contact: uwgchgs@u.washington.edu
Drafting Center: BCM
----- Project Information
Center project name: chr-3
Center clone name: RP11-7B12 (bc0603)
----- Summary Statistics

Sequencing vector: plasmid; 34% of reads
 Sequencing vector: unknown; 66% of reads
 Chemistry: Dye-terminator ET; 57% of reads
 Chemistry: Dye-terminator Big Dye; 43% of reads
 Assembly program: Phrap; version 0.930313
 Consensus quality: 152669 bases at least Q40
 Consensus quality: 152763 bases at least Q30
 Consensus quality: 152772 bases at least Q20
 Insert size: 152772; sum-of-contigs
 Quality coverage: 11.0x in Q20 bases; sum-of-contigs

Overlapping Sequences:

5': RPL11-7D10 (UWGC:bc0595) AC010959
 3': Mapping in progress

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
 all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

EcoRI				BglII			
SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt
8696	8700	1323	1309	12111	11940		
6	<800	6382	6519	2067	2160		
2532	2547	512	<800	8527	8714		
3139	3178	449	<800	1375	1342		
4975	5077	1133	1128	893	930		
2418	2547	2746	2762	4253	4526		
1931	1905	2037	2121	436	<800		
7381	7438	2845	2869	925	930		
1269	1212	3682	3728	2899	2789		
3691	3662	3956	3889	16122	16215		
3330	3376	1008	979	2045	2160		
8174	8159	5364	5343	3498	3433		

5231	5077	2323	2348	650	<800
2843	2906	1357	1309	13741	13713
5157	5077	5192	5120	60	<800
181	<800	10560	10611	5778	6013
1636	1608	758	<800	2421	2399
556	<800	10106	10165	1491	1472
1494	1474	5709	5695	2486	2399
360	<800	342	<800	1861	1904
5006	5077	2678	2762	12854	12717
638	<800	1893	1917	16	<800
5012	5077	304	<800	18	<800
256	<800	3190	3256	1286	1265
6568	6477	2488	2541	3630	3895
414	<800	5865	5895	1037	1057
2045	2079	1457	1442	4025	4293
17911	18009	5628	5695	3747	3895
756	<800	496	<800	6587	6827
5394	5422	886	979	2148	2160
1063	1053	4473	4390	10135	10124
1504	1474	1171	1128	1908	1904
3606	3376	465	<800	719	<800
4911	5077	7137	7196	2175	2160
208	<800	1039	979	2907	2789
5151	5077	234	<800	1064	1057
918	923	4865	4216	3648	3895
4060	4080	934	979	10740	10687
6	<800	6627	6519	2316	2399
3386	3376	1468	1442	3583	3433
675	<800	2168	2121	2782	2789
707	<800	951	979	188	<800
1393	1380	2339	2439	319	<800
8957	8700	2842	2869		
3442	3662	3748	3728		
4672	4584	1164	1128		
2212	2288	3610	3585		
4156	4080	3583	3585		

```

-----
1444      1474      457      <800
-----
7339      7136
-----
559      <800
-----
1883      1917
-----
8831      8854
-----
1515      1442
-----

```

FEATURES

```

source
1. .152772
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="3"

```

Query Match 73.3%; Score 22; DB 9; Length 152772;

Best Local Similarity 83.3%; Pred. No. 1.9e+02;

Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

QY 1 TTAATCATATCGTGTGTTGGTGTGTTG 30
    ||||| ||||| ||||| ||||| |||||
Db 34186 TTAATCATATCGTGTGTTGGTGTGTTG 34215

```

RESULT 6

AC010959/c

LOCUS 177672 bp DNA linear HTG 08-MAR-2000
 DEFINITION Homo sapiens clone RP11-7D10, WORKING DRAFT SEQUENCE, 25 unordered
 pieces.

ACCESSION AC010959

VERSION AC010959.3 GI:7209492

KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 177672)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens, clone RP11-7D10

Unpublished

2 (bases 1 to 177672)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,

Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouckgalter,B.,

Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,

Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,

Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,

Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,

McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,

Morrow,J., Naylor,J., Norton,C.H., O'Connor,T., O'Donnell,P.,

Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Tesfaye,S., Tirelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,

Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (28-SEP-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 8, 2000 this sequence version replaced gi:6978150.

All repeats were identified using RepeatMasker:

Snit, A.P.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L2922

FEATURES

source

1. .177672
 /organism="Homo sapiens"

Center clone name: 7 D.10

----- Summary Statistics

Sequencing vector: M13; M77815, 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 149787 bases at least Q40

Consensus quality: 162330 bases at least Q30

Consensus quality: 169351 bases at least Q20

Insert size: 182000; agarose-fp

Insert size: 175272; sum-of-contigs

Quality coverage: 3.9 in Q20 bases; agarose-fp

Quality coverage: 4.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 25 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

* 1 1129: contig of 1129 bp in length
* 1130 1229: gap of 100 bp
* 1230 2365: contig of 1136 bp in length
* 2366 2466: gap of 100 bp
* 2466 4110: contig of 1645 bp in length
* 4111 4210: gap of 100 bp
* 4211 5970: contig of 1760 bp in length
* 5971 6070: gap of 100 bp
* 6071 7611: contig of 1541 bp in length
* 7612 7711: gap of 100 bp
* 7712 9817: contig of 2066 bp in length
* 9818 12688: contig of 2871 bp in length
* 12689 16615: contig of 100 bp
* 16616 16715: gap of 100 bp
* 16716 20067: contig of 3352 bp in length
* 20068 23688: contig of 3521 bp in length
* 23689 27981: contig of 100 bp
* 27982 31601: contig of 4193 bp in length
* 31602 31701: gap of 100 bp
* 31702 37510: contig of 5809 bp in length
* 37511 42701: contig of 5091 bp in length
* 42702 42801: gap of 100 bp
* 42802 50237: contig of 7336 bp in length
* 50238 50138: gap of 100 bp
* 50138 57522: contig of 7285 bp in length
* 57523 66946: contig of 9324 bp in length
* 66947 74046: gap of 100 bp
* 74047 74537: contig of 7491 bp in length
* 74538 74637: gap of 100 bp
* 74638 78928: contig of 4291 bp in length
* 78929 79028: gap of 100 bp
* 79029 89116: contig of 10088 bp in length
* 89117 89216: gap of 100 bp
* 89217 98944: contig of 9728 bp in length
* 98945 99044: gap of 100 bp
* 99045 113038: contig of 13994 bp in length
* 113039 113138: gap of 100 bp
* 113139 131560: contig of 18422 bp in length
* 131561 131660: gap of 100 bp
* 131661 150982: contig of 19322 bp in length
* 150983 151082: gap of 100 bp
* 151083 177672: contig of 26590 bp in length.

```

```

/misc_feature /mol_type="genomic DNA"
/misc_feature /db_xref="taxon:9606"
/misc_feature /clone="RP11-7D10"
/misc_feature /clone_lib="RP11-11 Human Male BAC"
1. .1129
/misc_feature /note="assembly_fragment"
/misc_feature 1230. .2365
/misc_feature /note="assembly_fragment"
/misc_feature 2466. .4110
/misc_feature /note="assembly_fragment"
/misc_feature 4211. .5970
/misc_feature /note="assembly_fragment"
/misc_feature 6071. .7611
/misc_feature /note="assembly_fragment"
/misc_feature 7712. .3717
/misc_feature /note="assembly_fragment"
/misc_feature 9318. .12688
/misc_feature /note="assembly_fragment"
/misc_feature 12789. .16615
/misc_feature /note="assembly_fragment"
/misc_feature 16716. .20067
/misc_feature /note="assembly_fragment"
/misc_feature 20168. .23688
/misc_feature /note="assembly_fragment"
/misc_feature 23789. .27981
/misc_feature /note="assembly_fragment"
/misc_feature 28082. .31601
/misc_feature /note="assembly_fragment"
/misc_feature 31708. .37510
/misc_feature /note="assembly_fragment"
/misc_feature 37611. .42701
/misc_feature /note="assembly_fragment"
/misc_feature 42802. .50137
/misc_feature /note="assembly_fragment"
/misc_feature 50238. .57522
/misc_feature /note="assembly_fragment"
/misc_feature 57623. .66946
/misc_feature /note="assembly_fragment"
/misc_feature 67047. .74537
/misc_feature /note="assembly_fragment"
/misc_feature 74638. .78928
/misc_feature /note="assembly_fragment
clone_end:SP6
vector_side:left"
/misc_feature 79029. .89116
/misc_feature /note="assembly_fragment"
/misc_feature 89217. .98944
/misc_feature /note="assembly_fragment"
/misc_feature 99045. .113038
/misc_feature /note="assembly_fragment"
/misc_feature 113139. .131560
/misc_feature /note="assembly_fragment"
/misc_feature 131661. .150982
/misc_feature /note="assembly_fragment
clone_end:r7
vector_side:right"
/misc_feature 151083. .177672
/misc_feature /note="assembly_fragment"

```

ORIGIN

Query Match 73.3%; Score 22; DB 2; Length 177672;

Best Local Similarity 83.3%; Pred. No. 1.8e+02;

Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTGGTTATGTTG 30

||||| ||||| ||||| ||||| |||||

Db 96746 TTAATCATCTGTGTTTGGTTATGTTG 96717

RESULT 7

AC124915

LOCUS

DEFINITION Homo sapiens chromosome 3 clone RP11-717N7, complete sequence.

183748 bp

DNA

linear

PRI 23-APR-2003

AC124915
VERSION
AC124915.5 GI:30061451
HTG
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (Bases 1 to 183748)
AUTHORS
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
Haugen,E.D.
TITLE
Direct Submission
JOURNAL
Unpublished
REFERENCE
2 (Bases 1 to 183748)
AUTHORS
Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE
Direct Submission
JOURNAL
Submitted (19-JUN-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
REFERENCE
3 (Bases 1 to 183748)
AUTHORS
Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE
Direct Submission
JOURNAL
Submitted (13-NOV-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
REFERENCE
4 (Bases 1 to 183748)
AUTHORS
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
Haugen,E.D.
TITLE
Direct Submission
JOURNAL
Submitted (29-JAN-2003) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
REFERENCE
5 (Bases 1 to 183748)
AUTHORS
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
Haugen,E.D.
TITLE
Direct Submission
JOURNAL
Submitted (23-APR-2003) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
COMMENT
On Apr 23, 2003 this sequence version replaced gi:28014645.

COMMENT

Center: University of Washington Genome Center
Center Code: UMG
Web site: http://www.genome.washington.edu
Contact: umgctgs@u.washington.edu
----- Project Information
Center project name: chr-3
Center clone name: RP11-717N7 (bc0796)
----- Summary Statistics
Sequencing vector: plasmid; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 183705 bases at least Q40
Consensus quality: 183746 bases at least Q30
Consensus quality: 183748 bases at least Q20
Insert size: 183748; sum-of-contigs
Quality coverage: 8.0x in Q20 bases; sum-of-contigs

Overlapping Sequences:

5': RP11-444P10 (UMGC:bc0656) AC104445, 107092-bp overlap
3': RP11-7B12 (UMGC:bc0603) AC099326, 76535-bp overlap

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 207683)
Birren,B., Nussbaum,C. and Lander,E.
Mus musculus, clone RP23-348P7
Unpublished
2 (bases 1 to 207683)
Birren,B., Linton,L., Nussbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barria,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campoliano,A., Chang,J., Chazaro,B.,
Chospel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginsde,S., Gord,S., Coyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
Lander,T., Lebecky,J., Levine,R., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (31-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 207683)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barria,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Chospel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Fero,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Rafiez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Testaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-OCT-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 22, 2002 this sequence version replaced gi:20303870.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20P94
Center clone name: 348_P_7
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 204090 bases at least Q40
Consensus quality: 205414 bases at least Q30
Consensus quality: 206064 bases at least Q20
Insert size: 210000; agarose-fp

Insert size: 206683; sum-of-contigs
Quality coverage: 10.0 in Q20 bases; agarose-fp
Quality coverage: 10.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 2553: contig of 2553 bp in length
* 2554 2653: gap of 100 bp
* 2654 3493: contig of 840 bp in length
* 3494 3593: gap of 100 bp
* 3594 5369: contig of 1776 bp in length
* 5370 5469: gap of 100 bp
* 5470 7410: contig of 1941 bp in length
* 7411 7510: gap of 100 bp
* 7511 10014: contig of 2504 bp in length
* 10015 10114: gap of 100 bp
* 10115 18456: contig of 8342 bp in length
* 18457 18556: gap of 100 bp
* 18557 81827: contig of 63271 bp in length
* 81828 81927: gap of 100 bp
* 81928 114754: contig of 32827 bp in length
* 114755 114854: gap of 100 bp
* 114855 141736: contig of 26882 bp in length
* 141737 141836: gap of 100 bp
* 141837 189852: contig of 48016 bp in length
* 189853 189952: gap of 100 bp
* 189953 207683: contig of 17731 bp in length.

FEATURES
source

1..207683
/organism="Mus musculus"
/mol_type="Genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-348P7"
/clone_lib="RPCI-23 Female Mouse BAC"

misc_feature

1..2553
/note="assembly_fragment"
clone_end:SP6
vector_side:left

misc_feature

2654..3493
/note="assembly_fragment"

misc_feature

3594..5369
/note="assembly_fragment"

misc_feature

5470..7410
/note="assembly_fragment"

misc_feature

7511..10014
/note="assembly_fragment"

misc_feature

10115..18456
/note="assembly_fragment"

misc_feature

18557..81827
/note="assembly_fragment"

misc_feature

81928..114754
/note="assembly_fragment"

misc_feature

114855..141736
/note="assembly_fragment"

misc_feature

141837..189852
/note="assembly_fragment"

misc_feature

189953..207683
/note="assembly_fragment"
clone_end:T7
vector_side:right

ORIGIN

Query Match 73.3%; Score 22; DB 2; Length 207683;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

QY      1  TTAATCATATGCGTTTGGTTATGTTG 30
Db      123931 TTAATCATATGTTGTTTAAATATGTTG 123902

RESULT 9
BX248100/c      223489 bp      DNA      linear      VRT 22-JUL-2003
LOCUS      Zebrafish DNA sequence from clone CH211-222011 in linkage group 7,
DEFINITION      complete sequence.
ACCESSION      BX248100
VERSION      BX248100.9 GI:33146027
KEYWORDS      HTG.
SOURCE      Danio rerio (zebrafish)
ORGANISM      Danio rerio
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
              Cypriniformes; Cyprinidae; Danio.
REFERENCE      1 (bases 1 to 223489)
AUTHORS      Almeida, J.
TITLE      Direct Submission
JOURNAL      Submitted (22-JUL-2003) Wellcome Trust Sanger Institute, Hinxton,
              Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
              zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
              On Jul 22, 2003 this sequence version replaced gi:32168874.
              ----- Genome Center
              Center: Wellcome Trust Sanger Institute
              Direct code: SC
              Web site: http://www.sanger.ac.uk
              Contact: zfish-help@sanger.ac.uk
              -----
              During sequence assembly data is compared from overlapping clones.
              Where differences are found these are annotated as variations
              together with a note of the overlapping clone name. Note that the
              variation annotation may not be found in the sequence submission
              corresponding to the overlapping clone, as we submit sequences with
              only a small overlap as described above.
              This sequence was finished as follows unless otherwise noted: all
              regions were either double-stranded or sequenced with an alternate
              chemistry or covered by high quality data (i.e., phred quality >=
              30); an attempt was made to resolve all sequencing problems, such
              as compressions and repeats; all regions were covered by at least
              one plasmid subclone or more than one M13 subclone; and the
              assembly was confirmed by restriction digest, except on the rare
              occasion of the clone being a YAC.
              The following abbreviations are used to associate primary accession
              numbers given in the feature table with their source databases:
              Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
              on the WORMPEP database can be found at
              http://www.sanger.ac.uk/projects/C.elegans/wormpep/Clone-derived
              Zebrafish pUC subclones occasionally display inconsistency over the
              length of mononucleotide A/T runs and conserved TA repeats. Where
              this is found the longest good quality representation will be
              submitted.
              Repeat names beginning 'Dr' were identified by the Recon repeat
              discovery system (Zhihong Bao and Sean Eddy, submitted), and those
              beginning 'drr' were identified by Rick Waterman (Stephen Johnson
              lab, WashU). For further information see
              http://www.sanger.ac.uk/projects/D.rerio/fishmask.shtml
              CH211-222011 is from a CHORI-211 BAC library
              VECTOR: pTAR2AC2.1.
FEATURES      Location/Qualifiers
              source      1..223489
                        /organism="Danio rerio"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:7955"
                        /clone="CH211-222011"
                        /clone_lib="CHORI-211"
ORIGIN
Query Match      73.3%; Score 22; DB 5; Length 223489;
Best Local Similarity 83.3%; Pred.No. 1.8e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1  TTAATCATATGCGTTTGGTTATGTTG 30
Db      62379 TTAATCATATGCTTTTGTGTTTATTTG 62350

RESULT 10
BX649429/c      295094 bp      DNA      linear      HTG 20-OCT-2003
LOCUS      Danio rerio clone DKEY-1H4, WORKING DRAFT SEQUENCE, 7 unordered
DEFINITION      pieces.
ACCESSION      BX649429
VERSION      BX649429.3 GI:37776961
KEYWORDS      HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
ORGANISM      Danio rerio
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
              Cypriniformes; Cyprinidae; Danio.
REFERENCE      1 (bases 1 to 295094)
AUTHORS      McLaren, S.
TITLE      Direct Submission
JOURNAL      Submitted (18-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,
              Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
              zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
              On Oct 20, 2003 this sequence version replaced gi:36916794.
              ----- Genome Center
              Center: Wellcome Trust Sanger Institute
              Direct code: SC
              Web site: http://www.sanger.ac.uk
              Contact: zfish-help@sanger.ac.uk
              -----
              Assembly program: XGAP4; version 4.5
              Chemistry: Dye-terminator; 100% of reads
              Consensus quality: 290169 bases at least Q40
              Consensus quality: 290920 bases at least Q30
              Insert size: 294494; sum-of-contigs
              Insert size: 268382; 6.2% error; agarose-fp
              Quality coverage: 15.71x in Q20 bases; sum-of-contigs Quality
              coverage: 17.30x in Q20 bases; agarose-fp
              -----
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 7 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              * 1 35200: contig of 35200 bp in length
              * 35201 35300: gap of 100 bp
              * 35301 70426: contig of 35126 bp in length
              * 70427 70526: gap of 100 bp
              * 70527 152836: contig of 82310 bp in length
              * 152837 152836: gap of 100 bp
              * 152937 166852: contig of 13716 bp in length
              * 166853 166752: gap of 100 bp
              * 166753 188998: contig of 22246 bp in length
              * 188999 189098: gap of 100 bp
              * 189099 236930: contig of 47832 bp in length
              * 236931 237030: gap of 100 bp
              * 237031 295094: contig of 58064 bp in length.
              Location/Qualifiers
              1..295094
                /organism="Danio rerio"
                /mol_type="genomic DNA"
                /db_xref="taxon:7955"
                /clone="DKEY-1H4"
                /clone_lib="DanioKey"
              misc_feature      1..35200

```

```

/note="assembly_fragment:02678
fragment_chain:1"
misc_feature
3301..70426
/note="assembly_fragment:03573
fragment_chain:1"
misc_feature
70527..152836
/note="assembly_fragment:03044
fragment_chain:2"
152937..166652
/note="assembly_fragment:01565
fragment_chain:2"
166753..188998
/note="assembly_fragment:02056"
189099..236930
/note="assembly_fragment:03546"
237031..295094
/note="assembly_fragment:05006.0"
ORIGIN
Query Match 73.3%; Score 22; DB 2; Length 295094;
Best Local Similarity 83.3%; Pred. NO. 1.7e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTGGTTATGTTG 30
|||||
Db 121567 TTAATCATATGCTTTTGGTTATTTG 121538
|||||

RESULT 11
AX323523
LOCUS AX323523 6391 bp DNA linear PAT 07-JAN-2002
DEFINITION Sequence 11 from Patent WO0192565.
ACCESSION AX323523
VERSION AX323523.1 GI:18094271
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE
1 Olek,A.; Piepenbrock,C. and Berlin,K.
AUTHORS Olek,A.; Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with dna transcription
JOURNAL Patent: WO 0192563-A 11 06-DEC-2001;
Epigenomics AG (DE)
FEATURES
Location/Qualifiers
source
1..6391
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN
Query Match 72.0%; Score 21.6; DB 6; Length 6391;
Best Local Similarity 85.7%; Pred. NO. 4.3e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 TAATCATATGCGTTTGGTTATGTTT 29
|||||
Db 1000 TAATAATATGTGTTTGGATATGTT 1027
|||||

RESULT 12
AE002489/c
LOCUS AE002489 10099 bp DNA linear BCT 25-MAY-2000
DEFINITION Neisseria meningitidis serogroup B strain MC58 section 131 of 206
of the complete genome.
ACCESSION AE002489 AE002098
VERSION AE002489.1 GI:7226640
KEYWORDS
SOURCE Neisseria meningitidis MC58
ORGANISM Neisseria meningitidis MC58
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1 (bases 1 to 10099)

```

AUTHORS

Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C., Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F., Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D., Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D., Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E., Ciftone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H., Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V., Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R., Rappuoli,R. and Venter,J.C.

Direct Submission

Submitted (17-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers

1..10099

/organism="Neisseria meningitidis MC58"

/mol_type="genomic DNA"

/strain="MC58"

/db_xref="taxon:122586"

/note="serogroup: B"

complement (437..976)

/gene="NMB1401"

complement (437..976)

/gene="NMB1401"

/note="similar to GB.X59756 percent identity: 86.29; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="IS1016C2 transposase"

/protein_id="AAF41765.1"

/db_xref="GI:7226641"

translation="MKITHCKLKKVKOKELRFFVLVETARSADILGIHNSAVLFY RKIRVISHYALANAEVFGSVLDSYFGRRKRGKRGAGKVVVFGILKRGVY TVVDNAKSDTLPVVKQKIPDSIVYDLSSTYKLDVSGFIHRIHNSKFPADRO NHINGENFWQAKVLEK"

1313..1747

/gene="NMB1402"

1313..1747

/gene="NMB1402"

/note="hypothetical protein; identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF41766.1"

/db_xref="GI:7226642"

translation="MDNKKFNNLTNRYMYIPLVNLPLFPVQSQSPFIAGCL FALVKMSLDLFIQNHIVLNISAWADKKVFLIRIVVSWLAWMIWMCFISSSTW VCGAFCLNSELKIFRFGYSGSLYFLFILMDLNKLRSEI"

1788..1979

/gene="NMB1403"

1788..1979

/gene="NMB1403"

/note="FrpA/C-related protein; identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="FrpA/C-related protein"

/protein_id="AAF62333.1"

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Complete genome sequence of Neisseria meningitidis serogroup B strain MC58

Science 287 (5459), 1809-1815 (2000)

10710307

2 (bases 1 to 10099)

Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C., Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F., Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D., Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D., Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E., Ciftone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H., Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V., Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R., Rappuoli,R. and Venter,J.C.

Direct Submission

Submitted (17-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers

1..10099

/organism="Neisseria meningitidis MC58"

/mol_type="genomic DNA"

/strain="MC58"

/db_xref="taxon:122586"

/note="serogroup: B"

complement (437..976)

/gene="NMB1401"

complement (437..976)

/gene="NMB1401"

/note="similar to GB.X59756 percent identity: 86.29; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="IS1016C2 transposase"

/protein_id="AAF41765.1"

/db_xref="GI:7226641"

translation="MKITHCKLKKVKOKELRFFVLVETARSADILGIHNSAVLFY RKIRVISHYALANAEVFGSVLDSYFGRRKRGKRGAGKVVVFGILKRGVY TVVDNAKSDTLPVVKQKIPDSIVYDLSSTYKLDVSGFIHRIHNSKFPADRO NHINGENFWQAKVLEK"

1313..1747

/gene="NMB1402"

1313..1747

/gene="NMB1402"

/note="hypothetical protein; identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF41766.1"

/db_xref="GI:7226642"

translation="MDNKKFNNLTNRYMYIPLVNLPLFPVQSQSPFIAGCL FALVKMSLDLFIQNHIVLNISAWADKKVFLIRIVVSWLAWMIWMCFISSSTW VCGAFCLNSELKIFRFGYSGSLYFLFILMDLNKLRSEI"

1788..1979

/gene="NMB1403"

1788..1979

/gene="NMB1403"

/note="FrpA/C-related protein; identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="FrpA/C-related protein"

/protein_id="AAF62333.1"


```

/db_xref="GI:7413467"
/translation="MIGSGDTKQCKKFSACDGKHYHYDPLALDLDGDIETVTAKGFS
GSLKTERVNTMSIHSNPLN"
2014. .2742
/gene="NMB1404"
CDS
2014. .2742
/gene="NMB1404"
/notes="hypotheical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypotheical protein"
/protein_id="AAP41767.1"
/db_xref="GI:7226643"
/translation="MSHMKNNYLLVPIVLHIALVINIVFGVEVLEDFRAPLEPAN
VFLAVNLLPEKNIKXKLLPLPISIIIMVHISMINIKFKERHQIREQMISSITG
YAPHOSYFNVDNSGAKLQKUNRYRRETPYIDVVASOVNKSIRLSLVCGIHS
YAPCAFNPFAKPKVXIYFNQPGDFIDNVIFEINGNKSULLDKKYRTFLIENS
CIVILIILYKFNLLYRTYFNELE"
2756. .3958
/gene="NMB1405"
CDS
2756. .3958
/gene="NMB1405"
/notes="similar to GB:106302 PID:150255 SP:P55126 percent
identity: 93.58; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="PrpA/C-related protein"
/protein_id="AAP41768.1"
/db_xref="GI:7226644"
/translation="MSARSSALAEFGNMVANLVSAKNEKDISKRENYRQAGYSALLA
FNLASNIAPGTSSTHVNGTASVIAISLISGNISSAIQEHKDGKVINIRFQNILADL
YSLGGSLGTLIERGNQMGWGIPLAAGDIIAATATGDTGTISTEFPYFQNWKGP
GYELPDMRWYVWLPDQGNLWKELDNRSGQYHYDPLALDLDGDIETVAAGKFS
GSLFDHNGDRTATGWSADGCLLVRLDNGIILNGAELPDGNTKLAGDSFAKHGY
AAALDLSGNDNIINAADAFQSLRVQDLNQGISQANELRTLESLGQSLDLATKO
VFNKLGNGNTLAQQSYTTKDTGTAKMGDLLLLAADNLHGRFTNNKMLSIHVRENTISP
FVLGCLNK"
4389. .4751
/gene="NMB1406"
CDS
4389. .4751
/gene="NMB1406"
/notes="hypotheical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypotheical protein"
/protein_id="AAP41769.1"
/db_xref="GI:7226645"
/translation="MIAKSLFFRCQKIYFIYFILPCLVNLISYDGRIFYFIINPHT
LHICGILLVFCRIFFYENIPFTIFLNFISLFLIFLIFLIFLIFLIFLIFLIFLIF
LILPHVIFLILYKGRQI"
4748. .5877
/gene="NMB1407"
/notes="This region contains a gene with one or more
premature stops or framehifts, and is not the result of a
sequencing artifact; similar to GB:106302 PID:150255
SP:P55126 percent identity: 92.67; identified by sequence
similarity; putative"
5953. .6447
/gene="NMB1408"
CDS
5953. .6447
/gene="NMB1408"
/notes="hypotheical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypotheical protein"
/protein_id="AAP41770.1"
/db_xref="GI:7226646"
/translation="MNDHIVQIIRRFGLGRIFPYYSKSSIIIFSSYVYVYIYNQF
NYLSLIYLLPILCSIYMFIFFGKTKDTLTTRRRKKFFNSIFPLMIGSKKRL

```

```

GIGSYFLNLNLMIWCLMHRBQVPLNNLTLLLSFISLFFLLCDFVLLNLNVYVYFFK
LREA"
6575. .7003
/gene="NMB1409"
CDS
6575. .7003
/gene="NMB1409"
/notes="similar to GB:106302 PID:150255 SP:P55126 percent
identity: 77.78; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="PrpA/C-related protein"
/protein_id="AAP41771.1"
/db_xref="GI:7226647"
/translation="MSTSTYVAAHSGSLGKSVQBELQXTOSQLQVVRKMSKYNIN
NARVAELISLRKQMSQFSRKYNKNGSNWTLAQCGSYTKDGTGAQGDLLAADN
LHSRUTDKLISHVRENTISFVVGCLKQITINAYHLK"
7240. .7779
/gene="NMB1410"
CDS
7240. .7779
/gene="NMB1410"
/notes="hypotheical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypotheical protein"
/protein_id="AAP41772.1"
/db_xref="GI:7226648"
/translation="MDFLERIMSAFRKILLISCLLASCSFVETIFYMAISPEPVV
VDPLGKTKRSIELKQIKGRPYAISLGTNFHYDPKQGERWIDKLNYPNINISKIF
KVEDGKLLIDELUTERSKLGGVFGAGGKYSMHYIDFYLPGEYELFEISDSEYI
PLYDINNSIRIVVNARIQ"
complement (7950. .8603)
/gene="NMB1411"
CDS
complement (7950. .8603)
/gene="NMB1411"
/notes="similar to GB:X59756 percent identity: 84.00;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="IS1016C2 transposase"
/protein_id="AAP41773.1"
/db_xref="GI:7226649"

Query Match      72.0%; Score 21.6; DB 1; Length 10099;
Best Local Similarity 85.7%; Pred. No. 4e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 TAATCATATGCGTGTGTTGGTATGTT 29
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      6993 TAATCATATGCAITTAAGTTATTTGTT 6966

RESULT 13
AL513533/c
LOCUS      148151 bp DNA linear PRI 23-MAY-2002
DEFINITION Human DNA sequence from clone Rp11-341B24 on chromosome 10,
              complete sequence.
ACCESSION  AL513533
VERSION     AL513533.15 GI:21213137
KEYWORDS   HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 148151)
            Almeida, J.
            Direct Submission
            Submitted (23-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
            On May 25, 2002 this sequence version replaced gi:2038446.
            During sequence assembly data is compared from overlapping clones.
COMMENT

```

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, ENBL, Sw, SWISSPROT, Tr, TREML, Wp, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr10>
RP11-341B24 is from the library RPC1-11.2 constructed by the group of Pieper de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACE3.6.

FEATURES

Source

```
1..148151
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="10"
  /clone="RP11-341B24"
  /clone_lib="RPC1-11.2"
```

ORIGIN

```
Query Match      72.0%; Score 21.6; DB 9; Length 148151;
Best Local Similarity 85.7%; Pred. No. 2.7e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 2 TAATCATATGCGTTTGGTTATGTTG 29
|||||
Db 62859 TAATAATATGCGTTTGGTTTCTGTT 62832
```

RESULT 14

```
BX321886      169613 bp      DNA      linear      HTG 30-JUN-2003
LOCUS
```

```
DEFINITION    Danio rerio clone DKEYP-46C9, WORKING DRAFT SEQUENCE, 5 unordered
```

```
pieces
```

```
EX321886
```

```
ACCESSION    BX321886.5 GI:32398460
```

```
VERSION      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
```

```
KEYWORDS     Danio rerio (zebrafish)
```

```
SOURCE
```

```
ORGANISM     Danio rerio
```

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
```

```
Cypriniformes; Cyprinidae; Danio.
```

```
1 (bases 1 to 169613)
```

```
McLaren,S.
```

```
Direct Submission
```

```
Submitted (26-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
```

```
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
```

```
zfish-help@sanger.ac.uk Clone requests: clonesrequest@sanger.ac.uk
```

```
On Jul 1, 2003 this sequence version replaced gi:29603281.
```

```
----- Genome Center
```

```
Center: Wellcome Trust Sanger Institute
```

```
Center code: SC
```

```
Web site: http://www.sanger.ac.uk
```

```
Contact: zfish-help@sanger.ac.uk
```

```
----- Project Information
```

```
Center project name: zKp46C9
```

```
----- Summary Statistics
```

```
Assembly program: XGAP4; version 4.5
```

Chemistry: Dye-terminator; 100% of reads
Consensus quality: 167966 bases at least Q40
Consensus quality: 168229 bases at least Q30
Consensus quality: 168484 bases at least Q20
Insert size: 169213; sum-of-contigs
Insert size: 170871; 3.5% error; agarose-fp
Quality coverage: 8.69x in Q20 bases; sum-of-contigs Quality coverage: 8.68x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
* 1 4166: contig of 4166 bp in length
* 4167 4266: gap of 100 bp
* 4267 27503: contig of 23237 bp in length
* 27504 27603: gap of 100 bp
* 27604 85675: contig of 58072 bp in length
* 85676 85775: gap of 100 bp
* 85776 88119: contig of 2844 bp in length
* 88620 88719: gap of 100 bp
* 88720 169613: contig of 80894 bp in length.
```

FEATURES

Source

```
1..169613
  /organism="Danio rerio"
  /mol_type="genomic DNA"
  /db_xref="taxon:7955"
  /clone="DKEYP-46C9"
  /clone_lib="DanioKeyPilot"
  /clone_lib="DanioKeyPilot"
  /note="assembly fragment:00021"
  /fragment_chain:1
```

misc_feature

```
4267..27503
```

misc_feature

```
/note="assembly fragment:01353"
```

```
27604..85675
```

misc_feature

```
/note="assembly fragment:01439"
```

```
85776..88619
```

misc_feature

```
/note="assembly fragment:00679"
```

```
88720..169613
```

```
/note="assembly fragment:00756.0"
```

ORIGIN

```
Query Match      72.0%; Score 21.6; DB 2; Length 169613;
Best Local Similarity 85.7%; Pred. No. 2.6e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 3 AATCATATGCGTTTGGTTATGTTG 30
```

```
|||||
```

```
Db 27766 AATCATATGCGTTTGGTTATGTTG 27793
```

RESULT 15

```
AC144801/c
```

```
LOCUS
```

```
DEFINITION    Mus musculus chromosome 18 clone Rp24-276M2, complete sequence.
```

```
ACCESSION    AC144801
```

```
VERSION      AC144801.2 GI:38678656
```

```
KEYWORDS     HTG.
```

```
SOURCE      Mus musculus (house mouse)
```

```
ORGANISM     Mus musculus
```

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
1 (bases 1 to 174303)
```

```
WILSON,R.K.
```

```
TITLE The sequence of Mus musculus clone
```

```
JOURNAL Unpublished
```

```
REFERENCE 2 (bases 1 to 174303)
```

AUTHORS Wilson.R.K.
 TITLE Direct Submission
 JOURNAL Submitted (20-MAY-2003) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 REFERENCE 3 (bases 1 to 174303)
 AUTHORS Wilson.R.K.
 TITLE Direct Submission
 JOURNAL Submitted (04-DEC-2003) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 COMMENT On Dec 4, 2003 this sequence version replaced gi:30911155.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu
 Contact: submissions@watson.wustl.edu
 ----- Project Information -----
 Center project name: M_B50276M02

----- Location/Qualifiers -----
 1..174303
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="18"
 /clone="RP24-276M2"

ORIGIN

Query Match 72.0%; Score 21.6; DB 10; Length 174303;
 Best Local Similarity 85.7%; Pred. No. 2.6e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTTCGTTATGTGT 28
 |||||
 DB 64786 TCAATCATATGCGTCTGTTATGTGT 64759

Search completed: May 26, 2004, 17:21:57
 Job time : 425.308 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 15:20:44 ; Search time 389.487 Seconds
(without alignments)
3115.905 Million cell updates/sec

Title: US-10-676-299-4
Perfect score: 28
Sequence: 1 caacacataaccacaaagcatatgatt 28

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBank:
1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_cm.*
5: gb_cv.*
6: gb_pat.*
7: gb_ph.*
8: gb_pi.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_or.*
21: em_ov.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pi.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_hg_hum.*
31: em_hg_inv.*
32: em_hg_other.*
33: em_hg_mus.*
34: em_hg_pln.*
35: em_hg_rod.*
36: em_hg_mam.*
37: em_hg_vrt.*
38: em_sy.*
39: em_hgo_hum.*
40: em_hgo_mus.*
41: em_hgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	28	100.0	727	1	ECRFARSR	X16045 E. coli R-f
C 2	26.4	94.3	120826	1	AP005147	AP005147 Salmonell
C 3	21.6	77.1	169613	2	EX321886	EX321886 Danio rer
C 4	21.6	77.1	182051	9	AC007638	AC007638 Homo sapi
C 5	21.2	75.7	73000	2	AC090562	AC090562 Homo sapi
C 6	21.2	75.7	160811	9	AC020553	AC020553 Homo sapi
C 7	21.2	75.7	174303	10	AC144801	AC144801 Mus muscu
C 8	21.2	75.7	177109	2	AC118161	AC118161 Rattus no
C 9	21.2	75.7	184897	9	AC090919	AC090919 Homo sapi
C 10	21.2	75.7	193985	2	AC130151	AC130151 Rattus no
C 11	20.6	73.6	8391	6	AX323523	AX323523 Sequence
C 12	20.6	73.6	10099	1	AE002489	AE002489 Neisseria
C 13	20.6	73.6	65542	9	AC091766	AC091766 Homo sapi
C 14	20.6	73.6	110685	10	AL928957	AL928957 Mouse DNA
C 15	20.6	73.6	121636	2	AC122166	AC122166 Medicago
C 16	20.6	73.6	148151	9	AL513533	AL513533 Human DNA
C 17	20.6	73.6	160127	2	AC068977	AC068977 Homo sapi
C 18	20.6	73.6	166762	2	AC055852	AC055852 Homo sapi
C 19	20.6	73.6	198788	8	ATCHRI46	ATCHRI46 Arabidops
C 20	20.6	73.6	207674	8	ATPC8	ATPC8 Arabidops
C 21	20.6	73.6	217584	2	AC113299	AC113299 Mus muscu
C 22	20.6	73.6	218580	6	AC140927	AC140927 Mus muscu
C 23	20.6	73.6	349980	6	AX044032	AX044032 Sequence
C 24	20.2	72.1	34796	3	CF53011	CF53011 Caenorhabdi
C 25	20.2	72.1	151930	9	CNS01DVT	AL136038 Human chr
C 26	20.2	72.1	176345	2	AC132913	AC132913 Mus muscu
C 27	20.2	72.1	180044	2	AC130679	AC130679 Mus muscu
C 28	20.2	72.1	181902	2	AC026286	AC026286 Homo sapi
C 29	20.2	72.1	185702	2	AC138246	AC138246 Mus muscu
C 30	20.2	72.1	224453	2	AC134283	AC134283 Rattus no
C 31	20.2	72.1	230082	2	AC110303	AC110303 Rattus no
C 32	20.2	72.1	245066	2	AC125734	AC125734 Rattus no
C 33	20	71.4	1553	6	AX078799	AX078799 Sequence
C 34	20	71.4	1553	6	BD013422	BD013422 Transcrip
C 35	20	71.4	7276	8	AB089813	AB089813 Dauscus ca
C 36	20	71.4	13787	9	AE015654	AE015654 Shewanell
C 37	20	71.4	8957	9	AL136314	AL136314 Human DNA
C 38	20	71.4	105320	3	AC115680	AC115680 Dictyoste
C 39	20	71.4	108582	2	AC120459	AC120459 Homo sapi
C 40	20	71.4	119716	2	AC125480	AC125480 Medicago
C 41	20	71.4	131289	2	AC147278	AC147278 Pan trogl
C 42	20	71.4	137156	2	AC138846	AC138846 Homo sapi
C 43	20	71.4	138005	2	AC119008	AC119008 Rattus no
C 44	20	71.4	140410	2	AC139483	AC139483 Homo sapi
C 45	20	71.4	149030	9	AC005375	AC005375 Homo sapi

ALIGNMENTS

RESULT 1
ECRFARSR/c
LOCUS E. coli R-factor R773 arsr gene.
DEFINITION X16045
ACCESSION X16045
VERSION X16045.1 GI:42716
KEYWORDS arsenical resistance; arsr gene; Arsr protein; DNA-binding protein; regulatory protein; resistance gene.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 727)
AUTHORS San Francisco, M.J., Hope, C.L., Owolabi, J.B., Tisa, L.S. and Rosen, B.P.


```

MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
The transfer region of IncII plasmid R64: similarities between R64
tra and legionella icm/dot genes
Mol. Microbiol. 35 (6), 1348-1359 (2000)
20223621
PUBMED
REFERENCE
AUTHORS
TITLE
Organization and diversification of plasmid genomes: complete
nucleotide sequence of R64 genome
Unpublished
15 (bases 1 to 120826)
Komano,T., Sasaki,T., Tachibana,K., Furuya,N., Saito,Y.,
Suzuki,T. and Mizobuchi,K.
Sampei,G. and Komano,T., Sasaki,T., Tachibana,K., Furuya,N., Saito,Y.,
Suzuki,T., and Mizobuchi,K.
Direct Submission
Submitted (07-MAY-2002) Gen-ichi Sampei, The University of
Electro-Communications, Applied Physics and Chemistry; Chofugaoka
1-5-1, Chofu, Tokyo 182-8585, Japan (E-mail:sampeig@cc.uec.ac.jp,
Tel:81-424-43-5481, Fax:81-424-43-5501)
FEATURES
Source
1. 120826
/organism="Salmonella typhimurium"
/ecoli_type="genomic DNA"
/db_xref="taxon:602"
/lab_host="Escherichia coli strain K-12"
/plasmid="R64"
/notes="R64 strain drd-11"
/complement(join(20047..20418,3609..4147))
/notes="100 pct identical to sp:Y122_ECOLI(hypothetical
protein of insertion sequence IS2)
location complement(4143..4147) and location
complement(20047..20051) are duplicated"
378..467
/genes="repY"
378..467
/genes="repY"
/notes="100 pct identical to pir:A35445(RepY of plasmid
ColiB-P9)
possible regulatory reading frame"
/codon_start=1
/transl_table=11
/protein_id="BAB91567.1"
/db_xref="GI:20521503"
/translation="MKPYQRFNPVQCINTRHRSALSDSLNQV"
455..1486
/genes="repZ"
455..1486
/genes="repZ"
/notes="100 pct identical to sp:REPZ_ECOLI,sp:P18023(RepZ
of plasmid ColiB-P9)"
/codon_start=1
/transl_table=11
/protein_id="BAB91568.1"
/db_xref="GI:20521504"
/translation="KAGLQNTFYNAVHWSQLAPEQIRFWEDEYACRATTFILVEPERK
RTKRREGSTKPKCPNSWYPRVYKALGOLGHAYNELVKDPVTGEOUWRMSR
HPYVQKEFVGRKYAFREKORLLDAIWPVLVSPSDACTHTVGMSTVLAZEIHPKD
SGCHVPELVTSRLSLLAQVRFGLGVSEETWDEHQRQLRPYVWITPAGWQM
LGVWVKLEQQQQRURSEIRCOLIREGVLEDEDSVHAARKKWIYLRSDALKKR
REKAASKKAKKLLKPLFVDQQIYEMAYIRLKLPLPDBAYFCSDDDLKRLAIRLEQLQ
LTLAAPPPH"
2761..3114
/genes="arsR"
2761..3114
/genes="arsR"
/notes="100 pct identical to sp:ARR1_ECOLI,sp:P15905(arsR
of plasmid R773)"
/codon_start=1
/transl_table=11

```

```

/protein_id="BAB91569.1"
/db_xref="GI:20521505"
/translation="WLOLTPLQLPKNLSDETRLGIVLLRREMGELCVCDLCLMALDQSQ
PKISPHLMLRESGILLRKQKQVHYELSPHPSWAAQIIEQAWLSQDDDVQVIARK
LASVNCSSSKAVCI"
3162..3524
/genes="arsD"
3162..3524
/genes="arsD"
/notes="99 pct identical to sp:ARD1_ECOLI,sp:P46003(arsD of
plasmid R773)"
/codon_start=1
/transl_table=11
/protein_id="BAB91570.1"
/db_xref="GI:20521506"
/translation="MKTLMVDFPAMCCSTGVCCTDQALVDFSDVQVLMKQCGVOIE
RFNLAQQPMSPVQNEKVFATFASGAEGPLLLLDGTVMAGRYKPRARLARWFGIPL
DKVGLAPSGCCGNTSCC"
join(3542..3597,20833..22533)
/genes="arsA"
join(3542..3597,20833..22533)
/genes="arsA"
/notes="99 pct identical to pir:A25937(arsA of plasmid
R773, arsenical pump-driving ATPase)
location 3593..3597 and location 20833..20837 are
duplicated"
join(3598..4147,20047..20832)
/insertion_seq="IS2"
/complement(join(3609..4147,20047..20418))
/genes="yadA"
join(4148..7818,18208..20046)
/notes="different from Tn5393(gb:M963392) in IS1133 insert
position and terminal direct repeat"
Query Match 94.3%; Score 26.4; DB 1; Length 120826;
Best Local Similarity 96.4%; Pred.No. 1.3;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 CAACACATAACCAAAACGCATATGATT 28
|||||
DB 2707 CAACATATACCAAAACGCATATGATT 2680
|||||
RESULT 3
BX321886/c
LOCUS
DEFINITION
BX321886 169613 bp DNA linear HTG 30-JUN-2003
Danio rerio clone DKEXP-46C9, WORKING DRAFT SEQUENCE, 5 unordered
pieces
ACCESSION
BX321886
VERSION
BX321886.5 GI:32398460
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCES
Danio rerio (zebrafish)
ORGANISM
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE
1 (bases 1 to 169613)
AUTHORS
McLaren,S.
TITLE
Direct Submission
JOURNAL
Submitted (26-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jul 1, 2003 this sequence version replaced gi:29603281.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zkp46C9
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads

```

Consensus quality: 167966 bases at least Q40
 Consensus quality: 168229 bases at least Q30
 Consensus quality: 168484 bases at least Q20
 Insert size: 169213; sum-of-contigs
 Insert size: 170871; 3.5% error; agarose-ff
 Quality coverage: 8.69x in Q20 bases; sum-of-contigs Quality
 coverage: 8.68x in Q20 bases; agarose-ff

NOTE: This is a 'working draft' sequence. It currently
 consists of 5 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 4166: contig of 4166 bp in length
 4167 4266: gap of 100 bp
 4267 27503: contig of 23237 bp in length
 27504 27604: gap of 100 bp
 27604 85675: contig of 58072 bp in length
 85676 85775: gap of 100 bp
 85776 88619: contig of 2844 bp in length
 88620 88719: gap of 100 bp
 88720 169613: contig of 80894 bp in length.

FEATURES

Location/Qualifiers
 1..169613
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="DKY2-46C9"
 /clone_lib="DanioKeyPilot"
 1..4166
 /note="assembly_fragment:00021"
 /fragment_chain:1
 4267..27503
 /note="assembly_fragment:01353"
 /fragment_chain:1
 27604..85675
 /note="assembly_fragment:01439"
 /fragment_chain:1
 85776..88619
 /note="assembly_fragment:00679"
 88720..169613
 /note="assembly_fragment:00756.0"

ORIGIN

Query Match 77.1%; Score 21.6; DB 2: Length 169613;
 Best Local Similarity 85.7%; Pred. No. ie+02; 4; Indels 0; Gaps 0;
 Matches 24; Conservative 0; Mismatches 4

QY 1 CAACACATACCAACCAACGATATGATT 28
 Db 27793 CAACCATACCAACCAAGTCACATATGATT 27766

RESULT 4

AC007638/c
 LOCUS AC007638 182051 bp DNA linear PRI 02-DEC-2001
 DEFINITION Homo sapiens chromosome 17, clone RP11-515017, complete sequence.
 ACCESSION AC007638
 VERSION AC007638.8 GI:17226706
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 182051)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 17, clone RP11-515017
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 182051)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,

Baker,J., Baldwin,J., Barna,N., Beckerly,R., Berni,J., Brown,A.,
 Castle,A., Cerny,J., Colangelo,M., Collins,S., Collamore,A.,
 Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
 Donean,L., Doyle,M., Ferreira,B., FitzHugh,W., Forrest,C.,
 Funk,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
 Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
 Karakas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
 Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
 Maldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
 Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
 Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
 Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
 Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
 Direct Submission
 Submitted (22-MAY-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02144, USA
 3 (bases 1 to 182051)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chararo,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
 Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Govette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,K., Liu,G.,
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrim,J.,
 Meneses,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (02-DEC-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02144, USA
 On Dec 2, 2001 this sequence version replaced gi:13958524.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIRK
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: I458
 Center clone name: 515_O_17

FEATURES

Location/Qualifiers
 1..182051
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17"
 /clone="RP11-515017"
 /clone_lib="RP11-11 Human Male BAC"
 complement(1353..1468)
 /rpt_family="L2"
 complement(1805..2130)
 /rpt_family="AluY"
 2602..2760
 /rpt_family="AluYb"
 complement(3953..3656)
 /rpt_family="AluSx"

repeat_region
 repeat_region
 repeat_region
 repeat_region

```

repeat_region complement(4256..4412)
repeat_region /rpt_family="L2"
repeat_region 4584..6080
repeat_region /rpt_family="LIMEc"
repeat_region 6084..6439
repeat_region /rpt_family="AluSg"
repeat_region 6502..6762
repeat_region /rpt_family="AluSx"
repeat_region 6855..7007
repeat_region /rpt_family="LIMEc"
repeat_region 7008..7302
repeat_region /rpt_family="AluSg"
repeat_region 7303..7592
repeat_region /rpt_family="LIMEc"
repeat_region 7593..7697
repeat_region /rpt_family="MIR"
repeat_region 7694..7983
repeat_region /rpt_family="MIR3"
repeat_region 10083..10103
repeat_region /rpt_family="AT_rich"
repeat_region 11175..11198
repeat_region /rpt_family="AT_rich"
repeat_region 11941..12977
repeat_region /rpt_family="L2"
repeat_region 13421..13452
repeat_region /rpt_family="AT_rich"
repeat_region 13657..13961
repeat_region /rpt_family="AluSg"
repeat_region complement(14025..14134)
repeat_region /rpt_family="LIMEc"
repeat_region complement(14156..14355)
repeat_region /rpt_family="MIR"
repeat_region 15167..15343
repeat_region /rpt_family="MIR"
repeat_region complement(15613..15918)
repeat_region /rpt_family="AluSx"
repeat_region 15855..15885
repeat_region /note="PCR reads only"
repeat_region 15886..16038
repeat_region /note="PCR reads only"
repeat_region complement(16039..16103)
repeat_region /note="PCR reads only"
repeat_region complement(16252..16336)
repeat_region /rpt_family="L2"
repeat_region complement(16879..17191)
repeat_region /rpt_family="L2"
repeat_region complement(17385..17797)
repeat_region /rpt_family="MSTa"
repeat_region complement(18004..18306)
repeat_region /rpt_family="AluSg"
repeat_region complement(18406..18850)
repeat_region /rpt_family="L2"
repeat_region 18851..19257
repeat_region /rpt_family="MER31B"
repeat_region complement(19258..19308)
repeat_region /rpt_family="L2"
repeat_region complement(19543..20261)
repeat_region /rpt_family="LIMEc"
repeat_region 20409..20720
repeat_region /rpt_family="AluY"
repeat_region 20940..21074
repeat_region /rpt_family="MIR"
repeat_region 21287..21593
repeat_region /rpt_family="L2"
repeat_region complement(21594..21750)
repeat_region /rpt_family="FRAM"
repeat_region 21751..22415
repeat_region /rpt_family="L2"
repeat_region 22446..22805
repeat_region /rpt_family="AluSg/x"
repeat_region 22815..23063
repeat_region /rpt_family="L2"
repeat_region complement(23803..23918)
/rpt_family="MIR"
complement(25013..25306)
/rpt_family="AluJo"
complement(25483..25688)
/rpt_family="L2"
25769..26010
/rpt_family="MLTIL"
complement(26304..26420)
/rpt_family="MIR"
26831..26851
/rpt_family="AT_rich"
complement(26852..27023)
/rpt_family="FAM"
complement(27026..27326)
/rpt_family="AluJo"
complement(27337..27691)
/rpt_family="LIMEc"
27692..28054
/rpt_family="THE1B"
complement(28055..28551)
/rpt_family="LIMEc"
28601..28717
/rpt_family="MIR"
complement(29115..29419)
/rpt_family="AluSg"
complement(29420..29553)
/rpt_family="THE1B"
complement(29554..29913)
/rpt_family="THE1B"
complement(29914..31486)
/rpt_family="THE1B-int"
complement(31487..31680)
/rpt_family="THE1B"
complement(31681..31976)
/rpt_family="AluY"
complement(31977..32154)
/rpt_family="THE1B"
complement(33736..33840)
/rpt_family="MIR"

Query Match 77.1%; Score 21.6; DB 9; Length 182051;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CACACATACCAAAAGCATATGATT 28
Db 83655 CACTCATACCAAAAGCATATGAGT 83628

RESULT 5
AC090562 73000 bp DNA linear HTG 03-MAR-2001
LOCUS Homo sapiens chromosome 5 clone RP11-403N11 map 5, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
ACCESSION AC090562
VERSION AC090562.1 GI:13194292
KEYWORDS HTG; HTGS PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 73000)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 5, clone RP11-403N11
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 73000)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,I., Hulme,W., Iliev,I., Johnson,R.,

```


Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T.,
 Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P.,
 Marquis, N., Matthews, C., McCarthy, M., McSwan, P., McKernan, K.,
 McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mienga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
 O'Connor, F., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Ratta, R.,
 Rieack, M., Riley, R., Rise, C., Rogov, B., Roman, J., Rosetti, M.,
 Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
 Sougne, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
 Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
 Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
 Zembek, B., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smith, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L11595
 Center clone name: 403_N_11

NOTE: This record contains 89 individual
 sequencing reads that have not been assembled into
 contigs. Runs of N are used to separate the reads
 and the order in which they appear is completely
 arbitrary. Low-pass sequence sampling is useful for
 identifying clones that may be gene-rich and allows
 overlap relationships among clones to be deduced.
 However, it should not be assumed that this clone
 will be sequenced to completion. In the event that
 the record is updated, the accession number will
 be preserved.

1 710: contig of 710 bp in length
 711 810: gap of 100 bp
 811 1517: contig of 707 bp in length
 1518 1617: gap of 100 bp
 1618 2350: contig of 733 bp in length
 2351 2450: gap of 100 bp
 2451 3178: contig of 728 bp in length
 3179 3278: gap of 100 bp
 3279 3936: contig of 718 bp in length
 3937 4097: gap of 100 bp
 4098 4832: contig of 736 bp in length
 4833 4932: gap of 100 bp
 4933 5668: contig of 736 bp in length
 5669 5768: gap of 100 bp
 5769 6490: contig of 722 bp in length
 6491 6590: gap of 100 bp
 6591 7299: contig of 709 bp in length
 7300 7399: gap of 100 bp
 7400 8127: contig of 728 bp in length
 8128 8227: gap of 100 bp
 8228 8948: contig of 721 bp in length
 8949 9048: gap of 100 bp
 9049 9767: contig of 719 bp in length
 9768 9867: gap of 100 bp
 9868 10567: contig of 700 bp in length
 10568 10668: gap of 100 bp
 10669 11500: contig of 733 bp in length
 11501 12219: contig of 719 bp in length
 12220 12319: gap of 100 bp
 12320 13043: contig of 724 bp in length
 13044 13144: gap of 100 bp
 13145 13861: contig of 718 bp in length

13862 13961: gap of 100 bp
 13962 14678: contig of 717 bp in length
 14679 14778: gap of 100 bp
 14779 15505: contig of 727 bp in length
 15506 15605: gap of 100 bp
 15606 16313: contig of 708 bp in length
 16314 16413: gap of 100 bp
 16414 17117: contig of 704 bp in length
 17118 17217: gap of 100 bp
 17219 17943: contig of 726 bp in length
 17944 18043: gap of 100 bp
 18044 18750: contig of 707 bp in length
 18751 18850: gap of 100 bp
 18851 19591: contig of 741 bp in length
 19592 19691: gap of 100 bp
 19692 20414: contig of 723 bp in length
 20415 20514: gap of 100 bp
 20515 21241: contig of 727 bp in length
 21242 21341: gap of 100 bp
 21342 22055: contig of 714 bp in length
 22056 22155: gap of 100 bp
 22156 22873: contig of 718 bp in length
 22874 22973: gap of 100 bp
 22974 23735: contig of 762 bp in length
 23736 23835: gap of 100 bp
 23836 24553: contig of 718 bp in length
 24554 24653: gap of 100 bp
 24654 25367: contig of 714 bp in length
 25368 25467: gap of 100 bp
 25468 26183: contig of 716 bp in length
 26184 26283: gap of 100 bp
 26284 26997: contig of 714 bp in length
 26998 27097: gap of 100 bp
 27098 27824: contig of 727 bp in length
 27825 27924: gap of 100 bp
 27925 28654: contig of 730 bp in length
 28655 28754: gap of 100 bp
 28755 29464: contig of 710 bp in length
 29465 29564: gap of 100 bp
 29565 30297: contig of 733 bp in length
 30298 30397: gap of 100 bp
 30398 31123: contig of 728 bp in length
 31124 31223: gap of 100 bp
 31224 31955: contig of 732 bp in length
 31956 32055: gap of 100 bp
 32056 32778: contig of 723 bp in length
 32779 32878: gap of 100 bp
 32879 33582: contig of 704 bp in length
 33583 33682: gap of 100 bp
 33683 34371: contig of 689 bp in length
 34372 34471: gap of 100 bp
 34472 35194: contig of 723 bp in length
 35195 35294: gap of 100 bp
 35295 36008: contig of 714 bp in length
 36009 36108: gap of 100 bp
 36109 36809: contig of 701 bp in length
 36810 36909: gap of 100 bp
 36910 37647: contig of 738 bp in length
 37648 37747: gap of 100 bp
 37748 38459: contig of 712 bp in length
 38460 38559: gap of 100 bp
 38560 39301: contig of 742 bp in length
 39302 39401: gap of 100 bp
 39402 40140: contig of 739 bp in length
 40141 40240: gap of 100 bp
 40241 40946: contig of 706 bp in length
 40947 41046: gap of 100 bp
 41047 41774: contig of 728 bp in length
 41775 41874: gap of 100 bp
 41875 42602: contig of 728 bp in length
 42603 42702: gap of 100 bp
 42703 43404: contig of 702 bp in length
 43405 43504: gap of 100 bp

TITLE
 JOURNAL
 COMMENT

43505 44234: contig of 730 bp in length
 * 44235: gap of 100 bp
 * 44334: contig of 100 bp
 * 44335: contig of 725 bp in length
 * 45060 45159: gap of 100 bp
 * 45160 45877: contig of 718 bp in length
 * 45878 45977: gap of 100 bp
 * 45978 46710: contig of 733 bp in length
 * 46711 46810: gap of 100 bp
 * 46811 47539: contig of 729 bp in length
 * 47540 47639: gap of 100 bp
 * 47640 48358: contig of 719 bp in length
 * 48359 48458: gap of 100 bp
 * 48459 49186: contig of 728 bp in length
 * 49187 49286: gap of 100 bp
 * 49287 50010: contig of 724 bp in length
 * 50011 50110: gap of 100 bp
 * 50111 50818: contig of 708 bp in length
 * 50819 50918: gap of 100 bp
 * 50919 51621: contig of 703 bp in length
 * 51622 51721: gap of 100 bp
 * 51722 52431: contig of 710 bp in length
 * 52432 52531: gap of 100 bp
 * 52532 53240: contig of 709 bp in length
 * 53241 53340: gap of 100 bp
 * 53341 54057: contig of 717 bp in length
 * 54058 54157: gap of 100 bp
 * 54158 54887: contig of 730 bp in length
 * 54888 54987: gap of 100 bp
 * 54988 55712: contig of 725 bp in length
 * 55713 55812: gap of 100 bp
 * 55813 56542: contig of 730 bp in length
 * 56543 56642: gap of 100 bp

Query Match 75.7%; Score 21.2; DB 2; Length 73000;
 Best Local Similarity 88.5%; Pred. No. 1.8e+02;
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ACACATACCAACGACATATGATT 28
 |||||
 Db 21082 ACACATACCAACGACATATGATT 21107

RESULT 6
 AC020553/c
 LOCUS Homo sapiens BAC clone RP11-462D9 from 8, complete sequence.
 DEFINITION AC020553
 ACCESSION AC020553.4 GI:8570206
 VERSION HTG.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 160811)
 Sulston, J.E. and Waterston, R.
 Title Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074
 2 (bases 1 to 160811)
 Harkins, R., Hawkins, M., Drone, K. and Myers, M.
 Title The sequence of Homo sapiens BAC clone RP11-462D9
 JOURNAL Unpublished
 3 (bases 1 to 160811)
 Waterston, R.H.
 Title Direct Submission
 JOURNAL Submitted (03-JAN-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 4 (bases 1 to 160811)
 Waterston, R.H.
 Title Direct Submission
 JOURNAL Submitted (17-JUN-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 5 (bases 1 to 160811)
 Waterston, R.
 Title Direct Submission
 JOURNAL Submitted (07-OCT-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Jun 17, 2000 this sequence version replaced gi:7631006.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: sapiens@watson.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0462D09

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D.
 McPherson, Department of Genetics, Washington University, St. Louis
 MO. For additional information about the map position of this
 sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
 The RP11-11 human BAC library was made from the blood of one male
 donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, S.,
 Tateno, M., Catanese, J.J., and de Jong, P.J. (1998) An improved
 approach for construction of bacterial artificial chromosome
 libraries. Genomics 51:1-8. The clone may be obtained either from
 Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
 and coworkers at the Roswell Park Cancer Institute
 (http://bacpac.med.buffalo.edu)
 VECTOR: pBACe3.6
 NEIGHBORING SEQUENCE INFORMATION:
 Actual start of this clone is at base position 1 of RP11-462D9;
 actual end is at base position 160811 of RP11-462D9.

FEATURES
 Location/Qualifiers
 1..160811
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clone="RP11-462D9"
 /clone_lib="RP11-11"
 1014..1296
 /rpt_family="L1"
 1297..1991
 /rpt_family="L1"
 2004..3092
 /rpt_family="L1"
 4160..4474
 /rpt_family="MaLR"
 4475..4775
 /rpt_family="Alu"
 4776..4810
 /rpt_family="MaLR"
 6523..8335
 /rpt_family="L1"
 10561..10951
 repeat_region

repeat_region	/rpt_family="MaLR"
repeat_region	11062..11212
repeat_region	/rpt_family="MER53"
repeat_region	11269..11614
repeat_region	/rpt_family="BRVL"
repeat_region	11669..12155
repeat_region	/rpt_family="BRVL"
repeat_region	12204..12376
repeat_region	/rpt_family="BRVL"
repeat_region	12402..11363
repeat_region	/rpt_family="BRVL"
repeat_region	13205..13609
repeat_region	/rpt_family="BRVL"
repeat_region	13648..14000
repeat_region	/rpt_family="BRVL"
repeat_region	14144..14455
repeat_region	/rpt_family="Alu"
repeat_region	15093..15155
repeat_region	/rpt_family="BRVL"
repeat_region	15735..15825
repeat_region	/rpt_family="Alu"
repeat_region	17211..17378
repeat_region	/rpt_family="MIR"
repeat_region	17835..18247
repeat_region	/rpt_family="BRV1"
repeat_region	19073..19324
repeat_region	/rpt_family="L1"
repeat_region	19340..19642
repeat_region	/rpt_family="Alu"
repeat_region	19643..20579
repeat_region	/rpt_family="L1"
repeat_region	20667..20966
repeat_region	/rpt_family="Alu"
repeat_region	20967..21021
repeat_region	/rpt_family="L1"
repeat_region	21286..22180
repeat_region	/rpt_family="L1"
repeat_region	22557..22704
repeat_region	/rpt_family="L1"
repeat_region	22748..22868
repeat_region	/rpt_family="L2"
repeat_region	22868..23001
repeat_region	/rpt_family="L2"
repeat_region	23785..24443
repeat_region	/rpt_family="MaLR"
repeat_region	24144..24769
repeat_region	/rpt_family="BRV1"
repeat_region	25453..25566
repeat_region	/rpt_family="L2"
repeat_region	26889..27264
repeat_region	/rpt_family="L1"
repeat_region	27285..27773
repeat_region	/rpt_family="BRVK"
repeat_region	27774..27919
repeat_region	/rpt_family="L1"
repeat_region	28513..28669
repeat_region	/rpt_family="MIR"
repeat_region	29206..29374
repeat_region	/rpt_family="L2"
repeat_region	30889..30960
repeat_region	/rpt_family="MER2_type"
repeat_region	31006..31398
repeat_region	/rpt_family="BRV1"
repeat_region	31418..31621
repeat_region	/rpt_family="BRV1"
repeat_region	31645..31855
repeat_region	/rpt_family="MER2_type"
repeat_region	32671..32865
repeat_region	/rpt_family="L1"
repeat_region	33413..33575
repeat_region	/rpt_family="L1"
repeat_region	33813..34024
repeat_region	/rpt_family="BRV1"

```

repeat_region 34038. 34152
/rpt_family="L1"
repeat_region 35254. 36027
/rpt_family="ERV1"
repeat_region 36168. 36308
/rpt_family="L2"
repeat_region 36640. 37199
/rpt_family="ERV1"
repeat_region 38569. 38666
/rpt_family="L2"
repeat_region 39035. 39321
/rpt_family="MAIR"
repeat_region 39377. 39674
/rpt_family="L1"
repeat_region 39981. 40182
/rpt_family="ERV"
repeat_region 41247. 41845
/rpt_family="ERV1"
repeat_region 42018. 42166
/rpt_family="ERV1"
repeat_region 42294. 42655
/rpt_family="MAIR"
repeat_region 42855. 42916
/rpt_family="ERV1"
repeat_region 43220. 43446

```

```

Query Match      75.7%  Score 21.2;  DB 9;  Length 160811;
Best Local Similarity 88.5%  Pred. NO. 1.5e+02;
Matches 23;  Conservative 0;  Mismatches 3;  Indels 0;  Gaps 0;

Qy  3  ACACATAACCCAAAAACGCATATGATT 28
Db   75572  ACACATAACAAAAAGCTTATGATT 75547

```

RESULT 7	
LOCUS	AC144801
DEFINITION	Mus musculus chromosome 18 clone RP24-276M2, complete sequence.
ACCESSION	AC144801
VERSION	AC144801.2 GI:38678656
KEYWORDS	HTG.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 174303)
AUTHORS	Wilson, R.K.
TITLE	The sequence of Mus musculus clone
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 174303)
AUTHORS	Wilson, R.K.
TITLE	Direct Submission
JOURNAL	Submitted (20-MAY-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	3 (bases 1 to 174303)
AUTHORS	Wilson, R.K.
TITLE	Direct Submission
JOURNAL	Submitted (04-DEC-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	On Dec 4, 2003 this sequence version replaced gi:30311155.

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BB0276M02
-----
Location/Qualifiers
1. .174303
   /organism="Mus musculus"

```

```

/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="18"
/clone="RP24-276M2"

ORIGIN
Query Match      75.7%; Score 21.2; DB 10; Length 174303;
Best Local Similarity 88.5%; Pred. No. 1.5e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ACACATAACCAAAACGCGATATGATT 28
    |||||
Db 64759 ACACATAACCAAGCAGCGATATGATT 64784

RESULT 8
AC118161
LOCUS      Rattus norvegicus clone CH230-296C3, WORKING DRAFT SEQUENCE.
DEFINITION AC118161
ACCESSION  AC118161
VERSION    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS   Rattus norvegicus (Norway rat)
SOURCE     Rattus norvegicus
ORGANISM   Rattus

REFERENCE
AUTHORS    Muzny,D., Marie,, Metzger,M.Lee., Abramson,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyie,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroil,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dint,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganti,R., Garcia,A., Garner,T., Garza,M.,
            Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Haylak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Liu,J.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,F., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louleghed,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Polindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,P.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
            Valas,R., Vera,V., Villacana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,R., Warren,R., Wei,X., White,P.,
            Williams,G., Willson,R., Wleczky,R., Wooden,H., Worley,K.,

```

```

Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 177109)
Worley,K.C.
Direct Submission
Submitted (14-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 177109)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:22856412.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GVAU
Center clone name: CH230-296C3
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 172459 bases at least Q40
Consensus quality: 173621 bases at least Q30
Consensus quality: 174370 bases at least Q20
Estimated insert size: 177946; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 177109: contig of 177109 bp in length.
FEATURES
source
    1..177109
    /organism="Rattus norvegicus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10118"
    /clone="CH230-296C3"
    misc_feature
        1..1326
        /note="wgs_contig"
ORIGIN
Query Match      75.7%; Score 21.2; DB 2; Length 177109;
Best Local Similarity 88.5%; Pred. No. 1.5e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```



```

/rpt family=" (TC)n"
complement(7827..8025)
/rpt family="L1MED"
complement(8110..8282)
/rpt family="L2"
8283..9746
/rpt family="L1PA4"
complement(9747..9805)
/rpt family="L2"
complement(9938..9995)
/rpt family="MIR"
complement(10019..10368)
/rpt family="L1MA4"
10375..10463
/rpt family="MER5A"
10562..10988
/rpt family="AT_rich"
complement(11411..11476)
/rpt family="L2"
11721..12010
/rpt family="ALUJo"
12762..12891
/rpt family="L1MA8"
12944..13002
/rpt family="L1MA8"
13026..13785
/rpt family="L1ME3A"
14159..14262
/rpt family="ALUy"
complement(15527..15641)
/rpt family="L1PA6"
15642..15725
/rpt family="L1PA7"
16857..16885
/rpt family="AT_rich"
16927..17223
/rpt family="Tiger2"
complement(17232..17871)
/rpt family="L1MA"
complement(17878..18175)
/rpt family="L1MA"
complement(18164..24283)
/rpt family="L1PA5"
complement(23753..23764)
/notes="<30 qual SNGL region"
23776..23789
/notes="<30 qual SNGL region"
complement(23868..23870)
/notes="<30 qual SNGL region"
complement(23871..23875)
/notes="probably CCCC, possibly CCCCC"
complement(24351..25760)
/rpt family="L1M4c"
complement(25761..26052)
/rpt family="ALUSx"
complement(26053..26690)
/rpt family="L1M4c"
complement(26929..27128)
/rpt family="L1M4c"
27292..27316
/rpt family="AT_rich"

Query Match 75.7%; Score 21.2; DB 9; Length 184897;
Best Local Similarity 88.5%; Pred. No. 1.5e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ACACATAACCAAAACGCATATGATT 28
|||||
Db 109325 ACACATAACCAAAACGCATATGATT 109350
|||||

RESULT 10
AC130151/c

```

LOCUS AC130151 199985 bp DNA linear HTG 10-MAY-2003
 DEFINITION Rattus norvegicus clone CH230-42L16, *** SEQUENCING IN PROGRESS
 *** 6 unordered pieces.
 AC130151
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 VERSION Rattus norvegicus (Norway rat)
 KEYWORDS
 SOURCE
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 199985)
 Muzny, D.Marie., Mettler, M.Lee., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Cesari, H., Chen, A.,
 Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Y., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, I., Garza, M.,
 Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, J., King, B., Kovar, C.,
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensuhera, L., Loulseg, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwackelme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, P., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wlezyk, R., Woodden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G., and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 199985)
 Worley, K.C.
 Direct Submission
 Submitted (08-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 199985)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On May 10, 2003 this sequence version replaced gi:25009137.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GGUC
 Center clone name: CH230-42L16
 ----- Summary Statistics
 Assembly program: Atlas 3.0;
 Consensus quality: 130611 bases at least Q40
 Consensus quality: 136437 bases at least Q30
 Consensus quality: 139728 bases at least Q20
 Estimated insert size: 141175; sum-of-contigs estimation
 Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 170190: contig of 170190 bp in length
 * 170191 170230: gap of unknown length
 * 170291 192968: contig of 22678 bp in length
 * 192969 193068: gap of unknown length
 * 193069 194488: contig of 1420 bp in length
 * 194489 194588: gap of unknown length
 * 194589 195684: contig of 1096 bp in length
 * 195685 195784: gap of unknown length
 * 195785 197159: contig of 1385 bp in length
 * 197170 197269: gap of unknown length
 * 197270 199985: contig of 2716 bp in length.

FEATURES

source
 1..199985
 /organism="Rattus norvegicus"
 /mol_type="Genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-42L16"
 misc_feature 106671..108418
 /note="wgs contig"
 misc_feature 170291..172609
 /note="wgs_contig"
 ORIGIN
 Query Match 75.7%; Score 21.2; DB 2; Length 199985;
 Best Local Similarity 88.5%; Pred. No. 1.5e+02;
 Mismatches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 AACATATCAACCAAAACGCATATGATT 28
 |||||
 Db 180836 AACATATCAACCAACGCATATGATT 180811
 |||||

RESULT 11
 LOCUS AX323523/c 6391 bp DNA linear PAT 07-JAN-2002
 DEFINITION Sequence 11 from Patent WO0192565.
 ACCESSION AX323523
 VERSION AX323523.1 GI:18094271
 KEYWORDS synthetic construct
 SOURCE synthetic construct
 ORGANISM synthetic construct
 artificial sequences.
 REFERENCE 1
 AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
 TITLE Diagnosis of diseases associated with dna transcription
 JOURNAL Patent: WO 0192565-A 11 06-DEC-2001;
 Epigenomics AG (DE)
 FEATURES
 Location/Qualifiers
 source
 1..6391
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="chemically treated genomic DNA (Homo sapiens)"
 ORIGIN
 Query Match 73.6%; Score 20.6; DB 6; Length 6391;
 Best Local Similarity 85.2%; Pred. No. 4.8e+02;
 Mismatches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 AACCATATCAACCAAAACGCATATGATT 28
 |||||
 Db 1027 AACCATATCAACCAACATATATTATT 1001
 |||||
 RESULT 12
 LOCUS AE002489 10099 bp DNA linear BCT 25-MAY-2000
 DEFINITION Neisseria meningitidis serogroup B strain MC58 section 131 of 206
 of the complete genome.
 ACCESSION AE002489 AE002098
 VERSION AE002489.1 GI:7226640
 KEYWORDS
 SOURCE Neisseria meningitidis MC58
 ORGANISM Neisseria meningitidis MC58
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.
 REFERENCE 1 (bases 1 to 10099)
 AUTHORS Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
 Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F.,
 Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,
 Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
 Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,
 Ciftone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H.,
 Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V.,
 Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
 Rappuoli, R. and Venter, J.C.
 TITLE Complete genome sequence of Neisseria meningitidis serogroup B
 strain MC58
 JOURNAL Science 287 (5459), 1809-1815 (2000)
 MEDLINE 20175755
 PUBMED 10710307
 REFERENCE 2 (bases 1 to 10099)
 AUTHORS Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
 Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F.,
 Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,
 Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
 Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,
 Ciftone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H.,
 Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V.,
 Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
 Rappuoli, R. and Venter, J.C.
 TITLE Direct Submission
 JOURNAL Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 FEATURES
 Location/Qualifiers

```

source
1. 10099
/organism="Neisseria meningitidis MC58"
/mol_type="genomic DNA"
/strain="MC58"
/db_xref="taxon:122586"
/notes="serogroup: B"
complement(437. .976)
/gene="NMB1401"
CDS
/gene="NMB1401"
/notes="similar to GB:X59756 percent identity: 86.29;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="IS1016C2 transposase"
/protein_id="AAF41765.1"
/db_xref="GI:7226641"
/translations="MKITHCKLKKKKVOKELLRFVLEVTARSADILGHPSAVLFY
RKIRWTSYLAALANVEFGSVELDES YFGRRKGRGRGAAGKVVVFILKRGKV
YTVVDNAKSDTLMPVT KOKIMPDSIVYDLSLSYDKLDVSGVIHRIHNSKEFADRG
NHINGTENFWNAQKVLRK"
1313. .1747
/gene="NMB1402"
CDS
1313. .1747
/gene="NMB1402"
/notes="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF41766.1"
/db_xref="GI:7226642"
/translations="MDNKEFINLNRVYVYIPLVNLILFLPFYOSYSPFTALGCL
PALVRKXQSDPKLQNHVLLNLNKSADWKKVFLIRIVSWLAVMWIMWCFISESTW
VCGAFCLNSILSKIRGFGYSGSLYPLFILMDLNKIRESI"
1788. .1979
/gene="NMB1403"
CDS
1788. .1979
/gene="NMB1403"
/notes="FrpA/C-related protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="FrpA/C-related protein"
/protein_id="AAF62333.1"
/db_xref="GI:7413467"
/translations="MIGSGDTKQCKKFSACDGRYVYVYDPLALDLDGDIETVTKGFS
GSLKTERVNTMSIHSMLN"
2014. .2742
/gene="NMB1404"
CDS
2014. .2742
/gene="NMB1404"
/notes="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF41767.1"
/db_xref="GI:7226643"
/translations="MRMKKNLYLVFIVLHIALIVINIVPGFVFLPDPFAPLFPAN
VFLAVLPLEKLIKXKLLPLPISIIIMVTHISMINIKPKYFHEQIKRONISSITG
VTKPHDSVYVDSNCKYALKONHRYGRVIRETPDYDVASDVONKSIKLSLVCGIHS
YAPCANFIPAKPKVPIYFNQPGQFIDNVIPEINDGNKSLVLLDKYKTFPLENSV
CIVLILYLKFNLLLYRTYFNELE"
2756. .3958
/gene="NMB1405"
CDS
2756. .3958
/gene="NMB1405"
/notes="similar to GB:L06302 PID:L50255 SP:P55126 percent
identity: 93.58; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="FrpA/C-related protein"
/protein_id="AAF41768.1"
/db_xref="GI:7226644"
/translation="MSARSSALAEFCGMVANLVSAKNEKDISKREYVYKQAGYSALLA
FGNLASNIAPGSTSHIUNWASVTSPLSGNISSAIOEHKDGKNIKRPONILADL
YSLGLGSTLIEKGNMWSGPIPLAIGSLIIATAATATGTGTISTEEFYFNWKGFS
GIELFEDSRWYDMLPDGNNLWKELDNRSGQYHIYDEPLALDLDGDIETVAKGFS
GSLFDHNGIRIATGWSADDLVRLDNGNGIIDNGAELFGDNTKGDGSPAKHGYS
AALAEIDNSGDNRIIRAAAFQSLRVQDLNQDGI SQANELRTLELIGTOSLDLAYKD
VNKLNGNTLAQQGSYTKTGDGTARMGDLLLLAADNLHGRFTNKMLSISHVRENTISP
FVLTGCLNK"
4389. .4751
/gene="NMB1406"
CDS
4389. .4751
/gene="NMB1406"
/notes="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF41769.1"
/db_xref="GI:7226645"
/translations="MIKSLFFRCQKIYFIYFILFCYLYNLNYSYDGEIYFIYINFTH
LLICHGILLVFCRIFFYENIPFTIFLNFISLELIFLPIFTIRELIDSYIISINLF
LILIPHVIFLIYLGKQI"
4748. .5877
/gene="NMB1407"
/notes="This region contains a gene with one or more
premature stops or frameshifts, and is not the result of a
sequencing artifact; similar to GB:L06302 PID:L50255
SP:P55126 percent identity: 92.67; identified by sequence
similarity; putative"
5953. .6447
/gene="NMB1408"
CDS
5953. .6447
/gene="NMB1408"
/notes="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF41770.1"
/db_xref="GI:7226646"
/translations="NNDHIVQIRRFGLGRIFPYYSKSSIIIFGSVYVYIYNYQF
NYLSLLIYILPILCSIYMPFIPLGKTKDTLTERRKGFNSIFPRLIMIIGSEKKRL
GIGSFVLLNLLMIWCLMHRQVPLNNLTLLSLFISLFLLCDFVLLANVYVYFPK
LREA"
6575. .7003
/gene="NMB1409"
CDS
6575. .7003
/gene="NMB1409"
/notes="similar to GB:L06302 PID:L50255 SP:P55126 percent
identity: 77.78; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="FrpA/C-related protein"
/protein_id="AAF41771.1"
/db_xref="GI:7226647"
/translations="MSTOSTNYAAKHSIGKSVQRELOKTSQRLQVRVKMOSKYNIN
NKARVETSLRQMSQPSRYKVNKNGNSNTLAQGSYTKKDGTTAQAGDLLLLAADN
LHSTDKMLKLSHVRENTISPPVLGCLKQITINAYHLK"
7240. .7779
/gene="NMB1410"
CDS
7240. .7779
/gene="NMB1410"
/notes="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF41772.1"
/db_xref="GI:7226648"

```


Gene
CDS

/translation="MDPLIFITMSAFKILLIISGLLIASCSFVETIFVMAISPEPVV
VDFPKTKRISLKKIKGKPYALSITNFHYDPKQGESWIDDKLNPYNISVKIF
KVEBGRKLIIDELLTERSRKLGCGVFGAGGKYSNHIIDFPLPEGYLFEISDSBYI
PYDSINIRIVNARQ"
complement(7950..8603)
/gene="NM01411"
complement(7950..8603)
/gene="NM01411"
/note="similar to GB:X59756 percent identity: 84.00;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="IS1016C2 transposase"
/protein_id="AAF41773.1"
/db_xref="GI:7226649"

Query Match 73.6%; Score 20.6; DB 1; Length 10099;
Best Local Similarity 85.2%; Pred. No. 4.4e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AACACATACCAAAACGCATATGATT 28
|||||
Db 6966 AACAAATACCAATAATGCATATCAT 6992

RESULT 13
AC091766/c
LOCUS
DEFINITION Homo sapiens BAC clone RP11-393L14 from 7, complete sequence.
ACCESSION AC091766
VERSION AC091766.2 GI:15638835
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 65542)
AUTHORS Sulston, J.E. and Wilson, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PubMed 9847074
REFERENCE 2 (bases 1 to 65542)
AUTHORS Abbott, A. and Shah, N.
TITLE The sequence of Homo sapiens BAC clone RP11-393L14
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 65542)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 65542)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (18-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 65542)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 65542)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (29-APR-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 7 (bases 1 to 65542)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT

On Sep 18, 2001 this sequence version replaced gi:14269823.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: sapiens@wustl.edu
----- Summary Statistics

Center project name: H_NH0393L14

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/CTB/CHR7, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J., and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is CTB-22K14, 2000 bp overlap the
clone sequenced to the right is CTD-2009F14, 2000 bp overlap.
Actual start of this clone is at base position 54039 of CTB-22K14
actual end is at base position 9162 of CTD-2009F14.

FEATURES
source

Location/Qualifiers
1. 65542
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-393L14"
/clone_lib="RPCI-11"
316..432
/rpt_family="WMIR"
433..742
/rpt_family="Alu"
743..876
/rpt_family="WMIR"
936..962
/rpt_family="AT-rich"
1116..1547
/rpt_family="ERV1"
2053..2331
/rpt_family="Alu"
5389..5659
/rpt_family="Alu"
5672..5781
/rpt_family="ERV1"

```
repeat_region 6311. .6341
/rpt_family="AT_rich"
repeat_region 7230. .7310
/rpt_family="L2"
repeat_region 8289. .8317
/rpt_family="(TA)n"
repeat_region 8318. .8707
/rpt_family="ERV1"
repeat_region 8708. .8786
/rpt_family="(TA)n"
repeat_region 9152. .9277
/rpt_family="Mariner"
repeat_region 9375. .10146
/rpt_family="L1"
repeat_region 10354. .10384
/rpt_family="(T)n"
repeat_region 10646. .11085
/rpt_family="L2"
repeat_region 11119. .11166
/rpt_family="MIR"
repeat_region 11134. .11608
/rpt_family="L2"
repeat_region 11622. .12016
/rpt_family="MaLR"
repeat_region 12036. .12189
/rpt_family="L1"
repeat_region 12196. .12283
/rpt_family="CT-rich"
repeat_region 12364. .13100
/rpt_family="L1"
repeat_region 13162. .13767
/rpt_family="L1"
repeat_region 13772. .13988
/rpt_family="Alu"
repeat_region 14037. .14538
/rpt_family="L1"
repeat_region 14735. .15169
/rpt_family="L1"
repeat_region 15845. .16205
/rpt_family="L1"
repeat_region 16255. .16742
/rpt_family="L1"
repeat_region 16782. .16846
/rpt_family="AT_rich"
repeat_region 16885. .16952
/rpt_family="(TA)n"
repeat_region 17293. .17327
/rpt_family="(TCTCTG)n"
repeat_region 18033. .18411
/rpt_family="L1"
repeat_region 18530. .18569
/rpt_family="AT_rich"
repeat_region 18734. .19232
/rpt_family="L1"
repeat_region 19233. .20090
/rpt_family="L1"
repeat_region 20101. .20704
/rpt_family="L1"
repeat_region 20716. .21539
/rpt_family="L1"
repeat_region 22141. .22181
/rpt_family="CT-rich"
repeat_region 22226. .22870
/rpt_family="L2"
repeat_region 22973. .23007
/rpt_family="AT_rich"
repeat_region 23694. .23744
/rpt_family="MIR"
repeat_region 23745. .23943
/rpt_family="MER1_type"
repeat_region 23944. .24072
/rpt_family="MIR"
repeat_region 24096. .24130
```

```
repeat_region /rpt_family="AT_rich"
24412. .24630
/rpt_family="MaLR"
repeat_region 24620. .24811
/rpt_family="MaLR"
repeat_region 24812. .25124
/rpt_family="ERVL"
repeat_region 25391. .25428

Query Match 73.6%; Score 20.6; DB 9; Length 65542;
Best Local Similarity 85.2%; Pred. No. 3.1e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AACACATAACCAAAACCATATGATT 28
Db 64170 AACACATAACCAAAACCATATGATT 64144
```

RESULT 14

AL928957/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mouse DNA sequence from clone RP23-75E19 on chromosome 2, complete sequence.

AL928957.9 GI:38198241

HTG.

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 110685)

Wood, J.

Direct Submission

Submitted (05-NOV-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

On Nov 7, 2003 this sequence version replaced gi:30014178.

Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-75E19 is from the RPCI-23 Mouse BAC Library.

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6.

Location/Qualifiers

1. .110685

/organism="Mus musculus"

FEATURES

source

/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-75E19"
/clone_lib="RP23-75E19"

ORIGIN

Query Match 73.6%; Score 20.6; DB 10; Length 110685;
Best Local Similarity 85.2%; Pred. No. 2.8e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AACACATAACCAAAACGCAATGATT 28
|||||
DB 7324 AACAAACAACAAACAACTCATGATT 7298
|||||

RESULT 15

AC122166

LOCUS

AC122166 121636 bp DNA linear HTG 18-DEC-2003
Medicago truncatula clone mth2-33b23, WORKING DRAFT SEQUENCE, 5
unordered pieces.

ACCESSION

AC122166

VERSION

AC122166.22

KEYWORDS

HTG; HTGS PHASE1; HTGS DRAFT.

SOURCE

Medicago truncatula

ORGANISM

Medicago

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;

REFERENCE

1 (bases 1 to 121636)

AUTHORS

Shauli, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
Cook, D., Kim, D., and Roe, B.A.

TITLE

Medicago truncatula BAC Clone mth2-33b23

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 121636)

AUTHORS

Shauli, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
Cook, D., Kim, D., and Roe, B.A.

TITLE

Direct Submission

JOURNAL

Submitted (18-DEC-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

REFERENCE

3 (bases 1 to 121636)

AUTHORS

Shauli, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
Cook, D., Kim, D., and Roe, B.A.

TITLE

Direct Submission

JOURNAL

Submitted (18-DEC-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

COMMENT

On Dec 18, 2003 this sequence version replaced gi:39930621.

----- Genome Center

Center: Department Of Chemistry And Biochemistry

The University Of Oklahoma

Center code:UOKMOR

* NOTE: This is a 'working draft' sequence. It currently

* consists of 5 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 5209: contig of 5209 bp in length

* 5210 5308: gap of unknown length

* 5310 15708: contig of 10399 bp in length

* 15709 15808: gap of unknown length

* 15809 24261: contig of 8453 bp in length

* 24262 24361: gap of unknown length

* 24362 44670: contig of 20309 bp in length

* 44671 44770: gap of unknown length

* 44771 121636: contig of 76866 bp in length.

* Location/Qualifiers

FEATURES

source 1. 121636
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mth2-33b23"
/clone_lib="Medicago truncatula BAC library H2"

ORIGIN

Query Match 73.6%; Score 20.6; DB 2; Length 121636;
Best Local Similarity 85.2%; Pred. No. 2.8e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AACACATAACCAAAACGCAATGATT 28
|||||
DB 25003 AATACATATCCAAAGACATATGATT 25029
|||||

Search completed: May 26, 2004, 17:22:05
Job time : 397.487 secs

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 15:20:44 ; Search time 584.231 seconds
(without alignments)
3115.905 Million cell updates/sec

Title: US-10-676-299-5
Perfect score: 42
Sequence: 1 ctgcacttacacattcggtta.....tcatatgttttgactta 42

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_hgt_hum.*
- 31: em_hgt_inv.*
- 32: em_hgt_other.*
- 33: em_hgt_mus.*
- 34: em_hgt_pla.*
- 35: em_hgt_rod.*
- 36: em_hgt_mam.*
- 37: em_hgt_vrt.*
- 38: em_sy.*
- 39: em_hgtgo_hum.*
- 40: em_hgtgo_mus.*
- 41: em_hgtgo_other.*

Pred. No. is the number of results predicted by chance to have a

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	42	100.0	3492	1	ECARSRBC	X80057 E.coli gene
2	42	100.0	10240	1	AE000426	AE000426 Escherich
3	42	100.0	11524	1	AE015361	AE015361 Shigella
4	42	100.0	17941	2	AC145934	AC145934 Gallus ga
5	42	100.0	225419	1	ECOUW76	U00039 E. coli chr
6	42	100.0	242495	2	AC146183	AC146183 Pan trogi
7	42	100.0	289816	1	AE016992	AE016992 Shigella
8	37.2	88.6	11071	1	AE005575	AE005575 Escherich
9	37.2	88.6	267888	1	AP002565	AP002565 Escherich
10	33	78.6	301660	1	AE016768	AE016768 Escherich
11	25.6	61.0	90650	10	AF242431S2	AF242432 Mus muscu
12	25.6	61.0	198631	2	AC116741	AC116741 Mus muscu
13	25	59.5	88013	9	AC090511	AC090511 Homo sapi
14	25	59.5	150965	9	AC091915	AC091915 Homo sapi
15	25	59.5	169377	2	AC110578	AC110578 Homo sapi
16	25	59.5	178127	2	AC079076	AC079076 Homo sapi
17	25	59.5	198719	2	AC015716	AC015716 Homo sapi
18	24.6	58.6	149928	2	AC134950	AC134950 Danio rer
19	24.6	58.6	166112	2	AC104260	AC104260 Mus muscu
20	24.6	58.6	176496	10	AL772194	AL772194 Mouse DNA
21	24.6	58.6	196688	2	BX223087	BX223087 Danio rer
22	24.4	58.1	183699	2	BX255894	BX255894 Danio rer
23	24.4	58.1	221969	2	AC116134	AC116134 Mus muscu
24	24.2	57.6	156569	2	AC110190	AC110190 Homo sapi
25	24.2	57.6	164314	9	AC091691	AC091691 Homo sapi
26	24.2	57.6	194366	5	AL954179	AL954179 Zebrafish
27	24	57.1	447	11	BV006936	BV006936 sa99all.y
28	24	57.1	70329	2	AC087718	AC087718 Homo sapi
29	24	57.1	109891	9	AL353897	AL353897 Human DNA
30	24	57.1	162761	2	AC079916	AC079916 Homo sapi
31	24	57.1	198913	2	AC044915	AC044915 Homo sapi
32	24	57.1	207471	10	AC122249	AC122249 Mus muscu
33	24	57.1	212691	2	AC118594	AC118594 Mus muscu
34	24	57.1	272717	2	AC098469	AC098469 Rattus no
35	23.8	56.7	214795	2	AC110449	AC110449 Rattus no
36	23.8	56.7	233330	2	AC112548	AC112548 Rattus no
37	23.6	56.2	87548	5	BX323038	BX323038 Zebrafish
38	23.6	56.2	101904	5	AL935310	AL935310 Zebrafish
39	23.6	56.2	12100	8	AP005774	AP005774 Oryza sat
40	23.6	56.2	135070	2	AP003741	AP003741 Oryza sat
41	23.6	56.2	142711	10	AC121863	AC121863 Mus muscu
42	23.6	56.2	152433	2	AP004297	AP004297 Oryza sat
43	23.6	56.2	164936	5	BX255954	BX255954 Zebrafish
44	23.6	56.2	168990	2	AC122516	AC122516 Mus muscu
45	23.6	56.2	177205	2	AC079217	AC079217 Mus muscu

ALIGNMENTS

RESULT 1	ECARSRBC	3492 bp	DNA	linear	BCT 20-JUL-1995
LOCUS	E.coli genes arsrB, arsrC, arsrC.				
DEFINITION	X80057				
ACCESSION	X80057.1	GI:510824			
VERSION	arSR gene; arsrC gene; arsenate reductase; arsenic-efflux pump;				
KEYWORDS	arsenic-inducible repressor; arsr gene.				
SOURCE	Escherichia coli				
ORGANISM	Escherichia coli				
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
AUTHORS	Enterobacteriaceae; Escherichia.				
TITLE	Diorio, C., Cai, J., Marmor, J., Shinder, R. and DuBow, M.S.				
	An Escherichia coli chromosomal ars operon homolog is functional in				

```

arsenic detoxification and is conserved in gram-negative bacteria
J. Bacteriol. 177 (8), 2050-2056 (1995)
MEDLINE
PUBMED
7721697
REFERENCE
2 (bases 1 to 3492)
Diorio, C.
Direct Submission
Submitted (06-JUL-1994) C. Diorio, McGill University, 3775
University Street, Montreal, Quebec, H3A 2B4, CANADA
JOURNAL
Location/Qualifiers
FEATURES
source
1..3492
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:562"
repeat_region
528..538
/rpt_type=INVERTED
repeat_region
544..554
/rpt_type=INVERTED
gene
655..1061
/gene="arsR"
-35_signal
655..660
/gene="arsR"
-10_signal
678..683
/gene="arsR"
RBS
697..701
/gene="arsR"
CDS
708..1061
/gene="arsR"
/codon_start=1
/transl_table=1
/product="arsenic-inducible repressor"
/protein_id="CAA56362.1"
/db_xref="GI:516211"
/db_xref="GOA:P37310"
/translation="MSFLPIQLPKLADPRLTGIVLLSELGVCVCLTALDQSQ
PKISHLALLRESGILLRQKQKWHYRLSPHIPAWAKIIDEARCQEKVQATVRN
LARONCSGPSKNICS"
1071..2404
/gene="arsB"
-35_signal
1071..1076
/gene="arsB"
repeat_region
1072..1084
/rpt_type=INVERTED
repeat_region
1087..1099
/rpt_type=INVERTED
-10_signal
1088..1093
/gene="arsB"
RBS
1106..1110
/gene="arsB"
CDS
1115..2404
/gene="arsB"
/codon_start=1
/transl_table=1
/product="membrane-located arsenite efflux pump"
/protein_id="CAA56362.1"
/db_xref="GI:516211"
/db_xref="GOA:P37310"
/db_xref="GOA:P37310"
/translation="MLLAGAIFVLAIVLWQKGLGMSATLGAVALVGVVHPG
DIPVNNVWNTAFIAVIIISLLDESGFEWALHVSFNGRGRLLFTWVLIG
RAVAADFGDGLGALITPIVIAALLAGFSKGTLPVMAAGFIADTASLILVSNIV
NIVSADFGGLGFEVSNVMPVDIAIVATLVMLHYPKQIPQNDVALLKSPABAI
KDPATFTCWVLLALLGVFFVLEPLGIPVSAIAVAGLILEFVAKRGHAINTKVLR
GAPWQIVISLGMVLYVYGRNAGLTLYSLGVNLADNLAAATLGTGFTAFLSII
MNNPTLVGLAGSIDGSTAGSVIKAWVYAVIGDIPKIPITPGISLTLWLHLVISO
KNMTLSWGYFRTGIMTLPVLVFTLALALRLSFTL"
2417..2842
/gene="arsC"
-35_signal
2417..2842
/gene="arsC"
/codon_start=1

```

```

/transl_table=1
/product="arsenate reductase"
/protein_id="CAA56363.1"
/db_xref="GI:516212"
/db_xref="GOA:P37311"
/db_xref="SWISS-PROT:P37311"
/translation="MSNITIIYHNPAGTSRNTLEMRNSGTEPTIIHYLSTPPTRDEL
VKLIADMGISVALLRRGVPEYBELGAEDKFTDDRLLDFWLQFLINRPIVVTPLG
TRLCRSEVVLEILPDAQKGAFSKDEKGVVDEAGRLK"
2892..2905
/rpt_type=INVERTED
repeat_region
2910..2923
/rpt_type=INVERTED
ORIGIN
Query Match 100.0%; Score 42; DB 1; Length 3492;
Best Local Similarity 100.0%; Pred. No. 0.00044;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CTGCACATTACATTCGTTAGTCATATATATGTTTTCACCTTA 42
|||||
Db 621 CTGCACATTACATTCGTTAGTCATATATATGTTTTCACCTTA 662
RESULT 2
AE000426 10240 bp DNA linear BCT 01-DEC-2000
LOCUS Escherichia coli K12 MG1655 section 316 of 400 of the complete
DEFINITION Genome.
ACCESSION AE000426 U00096
VERSION AE000426.1 GI:1789910
KEYWORDS
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 (bases 1 to 10240)
Blattner, F.R.; Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)
97426617
PUBMED
9278503
REFERENCE
2 (bases 1 to 10240)
Blattner, F.R.
Direct Submission
AUTHORS
Blattner, F.R.
TITLE
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
JOURNAL
3 (bases 1 to 10240)
Blattner, F.R.
Direct Submission
AUTHORS
Blattner, F.R.
TITLE
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
JOURNAL
4 (bases 1 to 10240)
Blattner, F.R.
Direct Submission
AUTHORS
Blattner, F.R.
TITLE
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using GeneMark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG

```

Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

```

Location/Qualifiers
1. .10240
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
<1. .31
/notes="factor Sigma70; predicted +1 start at 3638369"
161. .1630
/genes="yhiP"
/notes="synonym: b3496"
161. .1630
/genes="yhiP"
/function="putative transport; Not classified"
/notes="0489; 99 pct identical amino acid sequence and
equal length to YH1P_ECOLI SW: P36837"
/codon_start=1
/transl_table=11
/product="putative transport protein"
/protein_id="AAC76521.1"
/db_xref="GI:1789911"
/translation="MNTTTPMGLQOPRFPMIFVFLWRFYGVGVQGLAVFFVKQ
LAFSQQAFVTFQAFALVYGLISIGYVGDHLLGTRIVLGLAVLAIGYPTGMSL
LRPDLFIAGLTAVGNLGFKANPALLSKYCPKPPKDRPDGAFLLPYMSINIGSIAL
SLAPVIAFRGYSVTYACAGLI IALLVIAICRGWKDISEPDPFPMFSKLAVL
LGSVMFVCMWEMHVENAVNLVLVLSIVVTIIPFOAFKDKTGERNKFVAVLML
EAVFVILVIAQMTSINFAINNHHIEILGFSINFSVFOALNPPFVVLASPILAGIYT
ELFISALGAMIALVPOHLMGFILGMWFLTGWFLITQAAFLGLGYVATFTAVPDNITDPLET
LPVYTVFGKIGLTVLGVAVVMLMVPMLKRMATPESH"
complement(1679. .2431)
/genes="yhiQ"
/notes="synonym: b3497"
complement(1679. .2431)
/genes="yhiQ"
/function="orf; Unknown"
/notes="f1250; 100 pct identical to 218 amino acids of
YH1Q_ECOLI SW: P37633 but has 32 additional N-ter
residues; gtg start, alternate starts possible"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC76522.1"
/db_xref="GI:1789912"
/translation="MKICLIDETGTGDSGLVLAARWGLEHDEDNLMALVLTPEHLEL
RKDEPKLGGIFVDYFGMAHRRKFGGRGEAVAKAVGKGDYLPDVVDATAGLGRD
AFVLASGVCEVRMLERNPVVALLDDGLAGYADAEIGWMLQELQLIHASSITALTDD
ITRPQVYVLDPMFPFHQKSAVKKENRQSVLSVFGDLDDGLLEPARLLATKRVVVK
RPDYAPPLANVATPVVTKGHRFDIYAGTPV"
complement(2439. .4481)
/genes="prlC"
/notes="synonym: b3498"
complement(2439. .4481)
/genes="prlC"

```

```

/EC_number="3.4.24.70"
/function="enzyme; Degradation of proteins, peptides,
glyco"
/notes="f680; 100 pct identical to OPDA_ECOLI SW: P27298;
CG Site No. 18031"
/codon_start=1
/transl_table=11
/product="oligopeptidase A"
/protein_id="AAC76523.1"
/db_xref="GI:1789913"
/translation="MTPPLLTPFELPPPSKILPSHVPAVTKALNDCRENVVVAQG
APYTWENICQLAEVDVVLGRIFFPVSHNSVKNPSPELBAVYQTLPLLSYSTWVGQ
HEGKYATRDURDGHYATLNTAKKAVDNALRDFETSGIGLPGKPEKQIGIEATRLS
ELGQYSNNVLDTATMGTKLVTDELAGMPEALAAAKAQAARKELEGYLLTLPIS
YLPWYTCDNALREEMYRAYSTRASDQGNAGKWSKVMEEILALREHLAQLAGFE
NYAFSKTKMAENPOQVLDELTLAKNARQGEKELAQALRAFAERFVDELQPMDI
AVYSKQKHLYSISDEOLRYPVPENKAVGLFEVVKRYIGITAKERKDVVWHEDVR
PELYDENELRGSPYLDLYARENKRGAWDDCVGOMRKADGSIQKQVAVLTCNFR
PVNGKPALTFDEVTITLFEHFGHGLHMLTETIAGVSGISGVPMDAVELESQFMENW
CWEPLAFLISGHYETGEPLKBLDMLAAKNYQAAFLILQLEFGGLDFRLHAEFR
PDQAKILETLAIKLVAVVPSWGRFFHAFSHIFAGGYAAGYSYLWADVLAAADA
PSRPEEGIFNRETQCSFLDNLIRGGSSEBMDLFRFRGREPQLDAMLEHYGK"
complement(2457. .2488)
/genes="prlC"
/notes="factor Sigma70; predicted +1 start at 3640781"
complement(2457. .2575)
/genes="prlC"
/notes="factor Sigma70; predicted +1 start at 3640869"
complement(2580. .2608)
/genes="prlC"
/notes="factor Sigma70; predicted +1 start at 3640904"
complement(4546. .4573)
/notes="factor Sigma32; predicted +1 start at 3642870"
4561. .4587
/notes="factor Sigma70; predicted +1 start at 3642925"
4684. .5526
/genes="yhiR"
/notes="synonym: b3499"
4684. .5526
/genes="yhiR"
/function="orf; Unknown"
/notes="c280; 100 pct identical amino acid sequence and
equal length to YH1R_ECOLI SW: P37634"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC76524.1"
/db_xref="GI:1789914"
/translation="MLSVRHSFHAGNHADVLKHTVOSLIIESLKDKRPFLLDTHAG
AGRYQIGSEHAERTGEYLEGIARIWQDDLPALAEAYINNVKHFNRSGQLRYPGSP
IARLLLRQDSQLQTLFELHPSDYPLRSFKDSRAKVEKADGFOQLKALPVPVRGL
ILIDPPYEMKTDYQAVSVGIAGYKRFATGIYALWVPVLRQIQIKRMHDLBATGIRK
ILQIELAVLPDSDRGMTASGMIVINPPFKLEQQMNVLPLWLSKLVPAQTGHATVSW
IVPG"
5291. .5303
/genes="yhiR"
/notes="central position to predicted promoter: -157.5"
/bound_moiety="Ada predicted site"
5421. .5449
/genes="yhiR"
/notes="factor Sigma70; predicted +1 start at 3643787"
5598. .6950
/genes="gor"
/notes="synonym: b3500"
5598. .6950
/genes="gor"
/EC_number="1.6.4.2"
/function="enzyme; Biosynthesis of cofactors, carriers:
thioredoxin, glutaredoxin, glutathione"
/notes="c450; 100 pct identical to GSHR_ECOLI SW: P06715;
CG Site No. 665"
/codon_start=1
/transl_table=11

```

/product="glutathione oxidoreductase"
 /protein_id="AAC76525.1"
 /db_xref="GI:1789915"
 /translation="MTKHYDYIAIGGGGGIASINRAAMYQKCALIEAKELGTCVN
 VCGPKVMHAAQIREALHMYGPDPDTINKENWETLIASRTAYIDRIHTSYENV
 LKXNVVDVKGFBFVDAKTLEVNGETITADHILLATGRSHPDIGVGYGIDSDGF
 FALPALPERVAVGAYIAVELAGVINGLAKTHLFPVRKHAPLRSFDPMTSETLVEVM
 1 CTTGCACTTACATTCGTTAAGTCATATATGTTTTTGACTTA 42
 7740 CTTGCACTTACATTCGTTAAGTCATATATGTTTTTGACTTA 7781

Query Match 100.0%; Score 42; DB 1; Length 10240;
 Best Local Similarity 100.0%; Pred. No. 0.00038;
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3
 AE015361
 LOCUS
 DEFINITION
 Shigella flexneri 2a str. 301 linear BCT 18-OCT-2002
 genome.
 ACCESSION
 AE015361
 VERSION
 AE015361.1
 KEYWORDS
 SOURCE
 ORGANISM

Shigella flexneri 2a str. 301
 Shigella flexneri 2a str. 301
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Shigella.

1 (bases 1 to 11524)
 Jin, Q., Yuan, Z.H., Xu, J.G., Wang, Y., Shen, Y., Lu, W.C., Wang, J.H.,
 Liu, E., Yang, J., Yang, F., Qu, D., Zhang, X.B., Zhang, J.Y., Yang, G.W.,
 Wu, H.T., Dong, J., Sun, L.L., Xue, Y., Zhao, A.L., Gao, Y.S., Zhu, J.P.,
 Kan, B., Chen, S.X., Yao, Z.J., He, B.K., Chen, R.S., Ma, D.L.,
 Qiang, B.Q., Wen, Y.M., Hou, Y.D. and Yu, J.

Genome sequence of Shigella flexneri 2a: insights into
 pathogenicity through comparison with genomes of Escherichia coli
 K12 and O157

Nucleic Acids Res. 30 (20), 4432-4441 (2002)
 12384590
 2 (bases 1 to 11524)
 Jin, Q., Shen, Y., Wang, J.H., Liu, H., Yang, J., Yang, F., Zhang, X.B.,
 Zhang, J.Y., Yang, G.W., Wu, H.T., Dong, J., Sun, L.L., Xue, Y.,
 Zhao, A.L., Gao, Y.S., Zhu, J.P., Chen, S.X., Yao, Z.J., Wang, J.,
 Lu, W.C., Qiang, B.Q., Wen, Y.M. and Hou, Y.D.

Direct Submission
 Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry
 of Public Health, 100 Yingxin Jie, Xuanwu Qu, Beijing 100052, P.R.
 China

FEATURES
 source
 1..11524
 /organism="Shigella flexneri 2a str. 301"
 /mol_type="genomic DNA"
 /strain="301"
 /serotype="2a"
 /db_xref="taxon:198214"
 67..1419
 /gene="gor"
 /locus_tag="SF3531"
 67..1419
 /gene="gor"
 /locus_tag="SF3531"
 /note="Residues 1 to 450 of 450 are 99 pct identical to
 residues 1 to 450 of a 450 aa protein from Escherichia
 coli O157:H7 ref: NP_312399.1"
 /codon_start=1
 /transl_table=1
 /product="glutathione oxidoreductase"
 /protein_id="AAC44987.1"
 /db_xref="GI:24053993"

/translation="MTKHYDYIAIGGGGGIASINRAAMYQKCALIEAKELGTCVN
 VCGPKVMHAAQIREALHMYGPDPDTINKENWETLIASRTAYIDRIHTSYENV
 LKXNVVDVKGFBFVDAKTLEVNGETITADHILLATGRSHPDIGVGYGIDSDGF

FALPALPERVAVGAGYIAVELAGVINGLAKTHLFPVRKHAPLRSFDPMTSETLVEVM
 NAEQPLHTNAIKRAVNNADGSLTLEDRSETVDCLIWAIIGREANDINLEAG
 VTKNEGYIVVDKYQNTNIECIYAGDNTGAVELTPVAAGRLSRLFNKPDDEHL
 DYSNIPTWFSHPPIGVGLTEPOAREYGDQVVKVYKSSFTAMVTATVTHRQPCRMK
 LVCVGPBEKIVIGIHGIFGMDMLQSFVALKMGATKKDPNTVAIHPTAAEEFVTR

complement(2332..2727)
 /gene="insB"
 /locus_tag="SF3532"
 complement(2332..2727)
 /gene="insB"

/locus_tag="SF3532"
 /note="Residues 1 to 131 of 131 are 100 pct identical to
 residues 37 to 167 of a 167 aa protein from Shigella
 flexneri pIR: T4475"
 /codon_start=1
 /transl_table=1

/product="orf, conserved hypothetical protein"
 /protein_id="AAN44988.1"
 /db_xref="GI:24053994"

/translation="MIVCASMDRHWGVGAKSORWLFPAYDRIIRRTVAHVFCERTL
 ATLERLSLSAFEVVVMVMTDGMPLYESRLKGLHVISKRYTORIERHNLNLROHLAR
 LVRKLSFSKSVLHDKVIGHYLNIRHYQ"
 complement(2754..3029)
 /gene="insA"
 /locus_tag="SF3533"
 complement(2754..3029)

/gene="insA"
 /locus_tag="SF3533"
 complement(2754..3029)

/locus_tag="SF3533"
 /note="Residues 1 to 91 of 91 are 98 pct identical to
 residues 1 to 91 of a 91 aa protein from Shigella flexneri
 ref: NP_085394.1"
 /codon_start=1
 /transl_table=1

/product="orf, conserved hypothetical protein"
 /protein_id="AAM44989.1"
 /db_xref="GI:24053995"

/translation="MASISIRPCSCATEGVRNGKSTAGHQRYLCSHCRTWLOQPT
 YTASQCHQKIIDMAMNGVGCASARIMGVGLNTVLRHLKNGSR"
 3071..3616

/locus_tag="SF3534"
 3071..3616
 /locus_tag="SF3534"
 /note="no database matches"
 /codon_start=1
 /transl_table=1

/product="orf, hypothetical protein"
 /protein_id="AAM44990.1"
 /db_xref="GI:24053996"

/translation="MAALPCYPSNCSKFLMKVWASPYCLANWESVASIALPSIS
 RSPRYPAIWHCCVAGYCWASKVSGFTAIRRI FQRRKRLMKPGDVMKRPQIS
 ATPDKTVPETVTFVAVNLAKHIWICTCVLSGRHYVTGRRYLCPDHRIGLAAERFR
 HRLECNARRGTGVYGRGSSG"
 3588..4772

/gene="arsB"
 /locus_tag="SF3535"
 3588..4772
 /gene="arsB"

/locus_tag="SF3535"
 /note="Residues 1 to 394 of 394 are 100 pct identical to
 residues 36 to 429 of a 429 aa protein from Escherichia
 coli O157:H7 EDL933 ref: NP_290073.1"
 /codon_start=1
 /transl_table=1

/product="arsenic pump membrane protein"
 /protein_id="AAN44991.1"
 /db_xref="GI:24053997"

/translation="MVTGVVHFGDIPVVMNIVMNAATAAFIAVIIISLLDESGFFWEA
 AFHVRWNGRGRLLFTVILGLAAVAALFANDGAALITPIVIMLLALGPKSGTTL
 ALVMAAGFTADTASLPLIVSNLVNIVSADFPGLGFRVAVMVPVDAIAIVATLVMHL
 LYFRKIDPNQDVALIKSPAEAKDPATFKTGWVLLLLNGPFLVPLGIPYSALAA
 VGLAILLPYKARGHAINTKVLKGAQWIVIFSLGMVLYVYGRNAGLTYLDSGLNV
 LADNGLMAATLGGTATLFLSSFWNNPVLVGLALSDGASGVIKEAMVYANVIC

DLGPKTPTIGSLATLWLHLVLSQNNMTISNGYFRTGIIIMTLPLVFLVTLAALRLSLP
TL"
4785..5210
/gene="arsc"
/locus tag="SF3536"
4785..5210
/gene="arsc"
/locus tag="SF3536"
/note="Residues 1 to 141 of 141 are 100 pct identical to
residues 1 to 141 of a 141 aa protein from Escherichia
coli K12 ref: NP_417960.1"
/codon start=1
/transl table=1
/product="arsenate reductase"
/protein id="AAN44992.1"
/db_xref="GI:24053998"
/translation="MNIITIHNPACGTGRNTLEMRNSGTEPTIHYLETPPTDEL
VKLIADMGISVRLAKNPEVBEGLAEDKPTDRLIDPMLQHPILINPIWTPPLG
TRLCRPSVVLTLPLDPAQGAFSKEDGKVDVDEAGRLK"
5560..6063
/gene="insB"
/locus tag="SF3537"
5560..6063
/gene="insB"
/locus tag="SF3537"
/note="Residues 1 to 167 of 167 are 100 pct identical to
residues 1 to 167 of a 167 aa protein from Shigella
flexneri pir: T44475"
/codon start=1
/transl table=1
/product="orf, conserved hypothetical protein"
/protein id="AAN44993.1"
/db_xref="GI:24053999"
/translation="MSRQCTHYGRWPOHGFTSLKLRPOSVTSRIOPGSDVIVCAEMD
EWGYGAKSRQWLFPYADRIRTVVAHVFGERTLATRELLLSLSAFEVVWMTDG
WPIYSLRGLKLVISKRYTORIERHNLRLQHLARLVKRSLSFSKSVLHDKVIGHY
LNKHVQ"
6088..7050
/locus tag="SF3538"
6088..7050
/locus tag="SF3538"
/note="Residues 34 to 319 of 320 are 34 pct identical to
residues 1 to 274 of a 274 aa protein from Campylobacter
jejuni ref: NP_282688.1"
/codon start=1
/transl table=1
/product="putative membrane protein"
/protein id="AAN44994.1"
/db_xref="GI:24054000"
/translation="MFVFLAVLSLPIVISAGVSLIRQKVPDHIQOMGARKGRGY
LAAALGVPFPCSTIPMLRGLLSAKAGPGPTTFLEVSLNPPIVGLMMWTFGW
KYLVAIIAAGVSLIASILDSLGFERHIIASKSSANCCAPAKTSPGTIYPIKVS
CCSPAALENPVVTCCTKAVVSNIPKILATKDALQPKDVPYLLGLVIGSTIYG
FIPSAIAAHAGADNPAPLPLSAVVGIPLYIRAEAVIPLASVLMTKMGLGALMALII
GSAGASLTVELLAKSMFRPMIAAFLTVILGVALMGYLTQLLF"
7631..8854
/locus tag="SF3539"
7631..8854
/locus tag="SF3539"
/note="Residues 1 to 365 of 407 are 75 pct identical to
residues 1 to 365 of a 371 aa protein from Escherichia
coli O157:H7 ref: NP_312403.1"
/codon start=1
/transl table=1
/product="orf, conserved hypothetical protein"
/protein id="AAN44995.1"
/db_xref="GI:24054001"
/translation="MSIDTPGI INTYHGDVYNCNTNTDNTVKTPTPKWPCDNEEQ
PINSIFSGVNPQFDAQOQLOINACHTNTYTNADYSKVVQVLYSLNNIETIS
SFELTOQTSILNQHNIRYKDKSVCCIIIVVPEFNNPIITKILVEEGIPERFSV
QTVSSDKNFAQADLPDIRDIQSLIKMKAUKIYKSNENRMEALAGCDFDNM
TQDMSKVLTSKSYFEDLLNFTDANLFTI FSHKENPIPKLHYEQHLDKQING
LFTSLITINDNSLRKAEIASRIIDFLKAVVNLSDFDILKTKQFKKICVYKQLEPT

Query Match 100.0%; Score 42; DB 1; Length 11524;
Best Local Similarity 100.0%; Pred. No. 0.00038;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CTGCACCTACACATCGTTAAGTCATATATATATGTTTGGACTTA 42
Db 2212 CTGCACCTACACATCGTTAAGTCATATATATATGTTTGGACTTA 2253
RESULT 4
LOCUS AC145934 179941 bp DNA linear HTG 01-AUG-2003
DEFINITION Gallus gallus chromosome UNK clone CH261-4211, *** SEQUENCING IN
PROGRESS ***, 44 unordered pieces.
AC145934
AC145934.1 GI:33386884
VERSION HTG: HTGS PHASE1.
KEYWORDS Gallus gallus (Chicken)
SOURCE Gallus gallus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 179941)
AUTHORS Wilson,R.K.
TITLE The sequence of Gallus gallus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 179941)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: J_AA042101
* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1176: contig of 1176 bp in length
* 1177 1276: gap of unknown length
* 1277 2558: contig of 1282 bp in length
* 2559 2658: gap of unknown length
* 2659 3848: contig of 1190 bp in length
* 3849 3948: gap of unknown length
* 3949 5266: contig of 1318 bp in length
* 5267 5367: gap of unknown length
* 5367 6402: contig of 1035 bp in length
* 6402 7599: contig of 1098 bp in length
* 7599 7600: gap of unknown length
* 7600 8831: contig of 1132 bp in length
* 8832 10481: contig of 1550 bp in length
* 10482 10581: gap of unknown length
* 10582 12037: contig of 1456 bp in length
* 12038 12137: gap of unknown length
* 12138 13596: contig of 1459 bp in length
* 13597 13696: gap of unknown length
* 13697 14911: contig of 1215 bp in length
* 14912 15011: gap of unknown length
* 15012 16279: contig of 1268 bp in length
* 16280 16379: gap of unknown length


```
* 16380 17948: contig of 1569 bp in length
* 17949 18048: gap of unknown length
* 18049 19612: contig of 1564 bp in length
* 19613 19712: gap of unknown length
* 19713 21723: contig of 2011 bp in length
* 21724 21823: gap of unknown length
* 21824 25266: contig of 3443 bp in length
* 25267 27105: gap of unknown length
* 27106 28536: contig of 1739 bp in length
* 28537 30905: gap of unknown length
* 30906 31005: contig of 1331 bp in length
* 31006 33224: gap of unknown length
* 33225 37384: contig of 4060 bp in length
* 37385 37484: gap of unknown length
* 37485 40797: contig of 3313 bp in length
* 40798 40897: gap of unknown length
* 40898 43117: contig of 2220 bp in length
* 43118 46243: contig of 3026 bp in length
* 46244 49765: contig of 3422 bp in length
* 49766 56038: gap of unknown length
* 56039 60189: gap of unknown length
* 60189 67006: contig of 6718 bp in length
* 67007 68878: contig of 1772 bp in length
* 68879 72905: gap of unknown length
* 72906 77513: contig of 4508 bp in length
* 77514 81177: gap of unknown length
* 81178 86612: contig of 5335 bp in length
* 86613 91514: gap of unknown length
* 91514 97538: gap of unknown length
* 97538 103312: contig of 5675 bp in length
* 103313 111608: gap of unknown length
* 111608 123016: contig of 11309 bp in length
* 123017 131760: contig of 8644 bp in length
* 131761 133264: gap of unknown length
* 133265 140148: gap of unknown length
* 140149 152045: contig of 11797 bp in length
* 152046 165190: gap of unknown length
* 165191 165290: contig of 13045 bp in length
* 165291 179941: gap of unknown length
* 179941 179941: contig of 14551 bp in length.
```

FEATURES

```
source
1..179941
/organism="Gallus gallus"
/mol_type="genomic DNA"
/db_xref="taxon:9031"
/chromosome="UNK"
/clone="CH261-4211"
misc_feature
1..1176
/feature="assembly_name:Contig10"
misc_feature
1277..2558
```

```
misc_feature
/feature="assembly_name:Contig17"
2659..3848
misc_feature
/feature="assembly_name:Contig25"
3949..5266
misc_feature
/feature="assembly_name:Contig30"
5367..6401
misc_feature
/feature="assembly_name:Contig31"
6502..7599
misc_feature
/feature="assembly_name:Contig32"
7700..8831
misc_feature
/feature="assembly_name:Contig36"
8932..10481
misc_feature
/feature="assembly_name:Contig39"
10582..12037
misc_feature
/feature="assembly_name:Contig40"
12138..13596
misc_feature
/feature="assembly_name:Contig41"
13697..14911
misc_feature
/feature="assembly_name:Contig42"
15012..16279
misc_feature
/feature="assembly_name:Contig43"
16380..17948
misc_feature
/feature="assembly_name:Contig44"
18049..19612
misc_feature
/feature="assembly_name:Contig45"
19713..21723
misc_feature
/feature="assembly_name:Contig46"
21824..25266
misc_feature
/feature="assembly_name:Contig47"
25367..27105
misc_feature
/feature="assembly_name:Contig48"
27206..28536
misc_feature
/feature="assembly_name:Contig49"
28637..30905
misc_feature
/feature="assembly_name:Contig50"
31006..33224
misc_feature
/feature="assembly_name:Contig51"
33325..37384
misc_feature
/feature="assembly_name:Contig52"
37485..40797
misc_feature
/feature="assembly_name:Contig53"
40898..43117
misc_feature
/feature="assembly_name:Contig54"
43218..46243
misc_feature
/feature="assembly_name:Contig55"
46344..49765
misc_feature
/feature="assembly_name:Contig56"
49866..56038
misc_feature
/feature="assembly_name:Contig57"
56139..60188
misc_feature
/feature="assembly_name:Contig58"
60289..67006
misc_feature
/feature="assembly_name:Contig59"
67107..68878
misc_feature
/feature="assembly_name:Contig6"
68979..72905
misc_feature
/feature="assembly_name:Contig60"
73006..77513
misc_feature
/feature="assembly_name:Contig61"
77614..81177
misc_feature
/feature="assembly_name:Contig62"
81278..86612
misc_feature
/feature="assembly_name:Contig63"
86713..91513
misc_feature
/feature="assembly_name:Contig64"
91614..97537
```

```
Query Match 100.0%; Score 42; DB 2; Length 179941;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 42, Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 CTGCCTTACATTCCTTAAGTCATATATGTTTTCACCTTA 42
|||||
```

```

Db      132301 CTGCACCTACACATTCGTTAAGTCATATATGTTTGGACTTA 132260

RESULT 5
LOCUS   EC00W76
DEFINITION 225419 bp DNA linear BCT 07-NOV-1996
ACCESSION U00039
VERSION   U00039.1 GI:466582
KEYWORDS
SOURCE   Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
  1 (bases 1 to 225419)
    Sofia,H.J., Burland,V., Daniels,D.L., Plunkett,G. III and
    Blattner,F.R.
  Analysis of the Escherichia coli genome. V. DNA sequence of the
  region from 76.0 to 81.5 minutes
  Nucleic Acids Res. 22 (13), 2576-2586 (1994)
  94316500
  PUBMED 8041620
REFERENCE 2 (bases 1 to 225419)
  Plunkett,G. III.
  Direct Submission
  Submitted (25-MAR-1994) Guy Plunkett III, Laboratory of Genetics,
  University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
  Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
  608-263-7459
  This sequence was determined as part of the E. coli Genome Project
  (Frederick R. Blattner, director) at the University of
  Wisconsin-Madison. Supported by award HG00301 from the NIH Human
  Genome Project. The entire sequence was independently determined
  from E. coli MG1655; overlaps and conflicts with other sequence
  determinations are annotated. The start of this entry overlaps the
  start of the entry EC00W82 (L10328) by 547 bp.

FEATURES
      source
      1..225419
        /organism="Escherichia coli"
        /mol_type="genomic DNA"
        /strain="K-12"
        /sub_strain="MG1655"
        /db_xref="taxon:562"
        /note="This sequence comprises the following lambda
        clones: EC19-98, EC27-273, EC17-131, EC30W5, EC30K609-3,
        EC27-334, EC30K606A-14, EC30K606A-12, EC30-154, EC27-385,
        EC27-1129, EC30W9, EC27-239, EC30W3, EC27-SF3955,
        EC18-200, EC27-1135, EC30R3900-8B, EC30R3900-5B, EC19W6,
        EC30K573-3; M13mp19 or Janus vectors were used for
        subcloning"
      misc_feature
        complement (<1..1221)
        /notes="corresponds to M28722; ECOGCT(928..2148)"
      gene
        complement (<1..1085)
        /gene="ggt"
        complement (<1..1085)
        /genes="ggt"
        /EC_number="2.3.2.2"
        /notes="CG Site No. 18280; cold-regulated"
      CDS
        /codon_start=1
        /transl_table=11
        /product="gamma-glutamyltranspeptidase"
        /protein_id="AAB18422.1"
        /db_xref="GI:466583"
        /translation="MIKPTFLRVAIAALLSGCSFAAAPPAPPVSYGVEDVPHYV
        RAKQGWAVDATATQGVDIILKEGNAVDAVAVYALAVTHPQNLGGGFWLIR
        SKNGWTTADPEMAPAKATRMELDDQNPDSKSLTSHLASGTPGTVAGPSLADK
        YCTMPLNKVQPAFLARDGFVNDALADDLATYGSVLPNHENSKAIFWKEGPELKK
        GDTLVQANLAKSLEMAENKGPDEFYKGTIAEQIAGMKNGKGLTKEDLALAYKAVET
        PISGDRYGVYSGMPSPSSGGHIHVQILNILENFDPMKYGFGSAMQIMAEAEKIAY
        ADRESEVLGDPDFVKVPWQALTNKVAKS"
      misc_feature
        <60..15171
        /notes="corresponds to lambda clone EC19-98"
      promoter
        1067..1094

```

```

/promoter
  complement (1085..1112)
  /note="39%"
  /note="43%"
  complement (1146..1173)
  /note="45%"
  complement (1181..2402)
  /note="corresponds to X56908; ECUGPQQ(1..1222)"
  1205..1645
  /note="alternate gene name yhhA"
  /codon_start=1
  /transl_table=11
  /label="ORF_0146a"
  /protein_id="AAB18423.1"
  /db_xref="GI:466584"
  /translation="MKRLLILTALIPFVFGAQPINTLNNPQGVQIPSQORMOTOMQ
  TQIQKGMALQKLTOTLQOQLENNINNSORVLOSQPGERNPARQMLPNTNGG
  MLNENPDSSLNQOQMLPERNGDMLNQPTPQDIPLKTGP"
  complement (1585..2383)
  /note="corresponds to X14437; ECUGPQ(1..819)"
  complement (1632..2375)
  /gene="ugpQ"
  complement (1632..2375)
  /gene="ugpQ"
  /EC_number="3.1.4.2"
  /codon_start=1
  /transl_table=11
  /product="glycerophosphocholine phosphodiesterase"
  /protein_id="AAB18424.1"
  /db_xref="GI:466585"
  /translation="MSNWPYPRIVAHGGGKLAPENTLASIDVGAKYGHMIEFDKAL
  SKDGEIFLHDNDLERTSGVAGELNQDILLRVDAGSWYSKMEKGEPLPLLSQVAE
  RCRGHWMANIEIKPTTGTGTLTKMVALAARELWAGMTPLSSFEIDALEAAQAA
  BELPGLLLDWRDDERELTLRLGCVSHLHKLNLKARVMQLKDAGLRLLVYTVNKP
  QRABLLKMGVDCICTDAIDVIGNFNTAQ"
  1693..1727
  /note="terminator-like sequence"
  2192..2233
  /note="Predicted bend of 73.99 degrees"
  complement (2225..6941)
  /note="corresponds to X13141; ECUGP(1..4717)"
  complement (2372..3481)
  /gene="ugpC"
  complement (2372..3481)
  /gene="ugpC"
  /function="sn-Glycerol-3-phosphate transport system"
  /notes="CG Site No. 17794"
  /codon_start=1
  /transl_table=11
  /protein_id="AAB18425.1"
  /db_xref="GI:912455"
  /translation="MPSCAAWSIVRNKMGKLAQVTKSMDGKTQVIKPLTLDVADGE
  FIVWPGCGCKTLLRMVAGLVRVTEGDIWINDQRTVTEMBPKDRGIAMVQNYALYP
  HMSVEENMAGLKIRGKQOIARVKEARILELDGLKRRPRLSGQQRQVAMGR
  ALVRDPVFLDPELSDAKLRVOMELRLOQLHRLKTTSLYVTHDQVEAMTLAQRV
  VMNGGVABQIGTPVVEYKPSLFSVAFSGSPAMNLLTGRVNNEGTHFELDGGIELP
  LKGGYRQYAGRKMTLIRPEHIALSSQAGGVPMVMDTLEILGADNLAHGWRGEQKLV
  VRLAQRPTAGSTLWLHLAENQLHLFDGETGQRY"
  complement (3444..4289)
  /gene="ugpB"
  complement (3444..4289)
  /gene="ugpB"
  /function="sn-Glycerol-3-phosphate transport system"
  /notes="CG Site No. 17791; membrane protein"
  /codon_start=1
  /transl_table=11
  /protein_id="AAB18426.1"
  /db_xref="GI:466587"
  /translation="MIENRPMLTIFSHMTLILGLAVILPPLVAVFAATLDKQAVAA
  PMTILIPGHLLENLHNIVWGVGNSAPFFWRMLNSFNAPSITLTKITVSMLSAFAL
  VWFFPPLNLFWMIFITLMLPVEIRPPTVEIRLQMLDSYAGLTLPNLSATATP
  LFRQFFMTLPDELVEAARIDGASPMRPFCDIVFPLSKTLAALFITPITVGNQYILMP
  LLIIITDVLGTVAGIKGMIATGCTTWNVMVAMLLTLPVPPVILVMQRAFVRGL
  VDSEK"

```


Query Match 100.0%; Score 42; DB 2; Length 242495;
 Best Local Similarity 100.0%; Pred. NO. 0.00026;
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCATTACACATTCGTTAAGTCATATATGTTTGGACTTA 42

Db 3312 CTGCATTACACATTCGTTAAGTCATATATGTTTGGACTTA 3353

RESULT 7

AE016992/c 289816 bp DNA linear BCT 22-APR-2003
 LOCUS Shigella flexneri 2a str. 2457T section 15 of 16 of the complete
 DEFINITION genome.

ACCESSION AE016992

VERSION AE014073

KEYWORDS AE016992.1 GI:30043426

SOURCE

ORGANISM Shigella flexneri 2a str. 2457T
 Shigella flexneri 2a str. 2457T
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Shigella.

REFERENCE

AUTHORS Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W.,
 Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A.,
 Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S.,
 Schwartz, D.C. and Blattner, F.R.

TITLE Complete Genome Sequence and Comparative Genomics of Shigella

flexneri Serotype 2a Strain 2457T

Infect. Immun. 71 (5), 2775-2786 (2003)

JOURNAL

PUBMED 12704152

REFERENCE

AUTHORS Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W.,
 Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A.,
 Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S.,
 Schwartz, D.C. and Blattner, F.R.

TITLE Direct Submission

JOURNAL Submitted (13-JUN-2002) Genetics Laboratory, University of

Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA

FEATURES

source

1. 289816

/organism="Shigella flexneri 2a str. 2457T"

/mol_type="genomic DNA"

/strain="2457T"

/serotype="2a"

/db_xref="taxon:198215"

complement (238..450)

/gene="cspA"

/locus_tag="S4179"

complement (238..450)

/gene="cspA"

/locus_tag="S4179"

/function="regulator; Adaptations, atypical conditions"

/note="residues 1 to 70 of 70 are 100.00 pct identical to

residues 1 to 70 of 70 from Escherichia coli K-12 : B3556"

/codon_start=1

/transl_table=11

/product="cold shock protein 7.4, transcriptional

activator of hns"

/protein_id="AAP19147.1"

/db_xref="GI:30043427"

/translation="MSGQMTGVVXWFRNADKGFITPDGSKDVVFHFSATQNDGYKS

LDEQKVSFTTESAKGPAAGNVTSL"

complement (896..1021)

/gene="yiaC"

/locus_tag="S4180"

complement (896..1021)

/gene="yiaC"

/locus_tag="S4180"

/note="residues 1 to 41 of 41 are 100.00 pct identical to

residues 1 to 41 of 96 from Escherichia coli K-12 : B3555"

/codon_start=1

/transl_table=11

gene

CDS

/product="hypothetical protein"
 /protein_id="AAP19148.1"
 /db_xref="GI:30043428"
 /translation="MEYKDPMHLLSLLEQIVPKDTQKIFLTHRTTCTSEIQL"
 1335..2165
 /gene="yiaP"
 /locus_tag="S4181"
 1335..2165
 /gene="yiaP"
 /locus_tag="S4181"
 /note="residues 1 to 276 of 276 are 94.56 pct identical to
 residues 1 to 276 of 276 from Escherichia coli K-12 :
 B3554"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAP19149.1"
 /db_xref="GI:30043429"
 /translation="MILPGLRRKGILOACPGLSRQTRVCRCLFLGERSKKMWTG
 KSSRWPAFLAALIMVSLSCFTKEGDKAFIDFLQNTVMRSGBRLFTLTADQKKQ
 FGPFVSYAILYGVQOVNOMDGLRPVDSVNAIRPODYVTSQGLPRMNGSLGV
 LAQOLNKLQADAAHSALKOSDILKEPFDQATKVTYTPADALQPLIPAAQTFTQOL
 VMGDDYTAQQTQVSFVANGIQFTTSQASSEYKLLIAPLPAQHQAFNQATTAATATC
 complement (2215..3189)
 /gene="yiaE"
 /locus_tag="S4182"
 complement (2215..3189)
 /gene="yiaE"
 /locus_tag="S4182"
 /function="putative enzyme; Not classified"
 /note="residues 1 to 324 of 324 are 94.13 pct identical to
 residues 5 to 328 of 328 from Escherichia coli K-12 :
 B3553"
 /codon_start=1
 /transl_table=11
 /product="putative dehydrogenase"
 /protein_id="AAP19150.1"
 /db_xref="GI:30043430"
 /translation="MPEVILYKALPDDLQRLQAHFTVHVQANLSPTVVQNAIPA
 EAELGSGNENVDALILEKPKLRATSTISGYDNFDVDALTKAKILMHTPTVLITET
 VADTLMALVLTARRVVEAERVKAGEWTASIGDWTGTDVHHKTLGIVGMRIGMAL
 AQRVHGFNMPILYNARGHKBEARNARYCDLTLOESDFVCLLPLTDETHLUF
 GAQFAKMSAIFINARGGPVVDENALIAALOKGEIHAAGLDVPEGEPLSVDSFLLS
 MANVAVPHIGSAITHETRYGMAACAVNDLIDALQGVKKNVPHVAD"
 complement (3293..3919)
 /gene="yiaD"
 /locus_tag="S4183"
 complement (3293..3919)
 /gene="yiaD"
 /locus_tag="S4183"
 /function="putative membrane; Not classified"
 /note="residues 1 to 208 of 208 are 88.94 pct identical to
 residues 12 to 219 of 219 from Escherichia coli K-12 :
 B3552"
 /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein"
 /protein_id="AAP19151.1"
 /db_xref="GI:30043431"
 /translation="MSGALVSGCTTNPYTGEEAGKSAIGAGLSLVGAGIGALSSS
 KDRGKALITGAAGALGGVGYMDVQEAKLDRMRGTGVSVTRSGDNIILNPNP
 VTFDSSAPLTPAGANTLTGVAAMLKEYPKTAVNVITGYTDTGCHDLNMLSLQORADS
 VASALITQGVDSRIRITQGLGPANPIASNTABGKAQNRREITLSPL"
 4160..6439
 /gene="bisC"
 /locus_tag="S4184"
 4160..6439
 /gene="bisC"
 /locus_tag="S4184"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 Biotin"
 /note="residues 21 to 759 of 759 are 98.78 pct identical

gene

CDS

/product="hypothetical protein"
 /protein_id="AAP19150.1"
 /db_xref="GI:30043430"
 /translation="MPEVILYKALPDDLQRLQAHFTVHVQANLSPTVVQNAIPA
 EAELGSGNENVDALILEKPKLRATSTISGYDNFDVDALTKAKILMHTPTVLITET
 VADTLMALVLTARRVVEAERVKAGEWTASIGDWTGTDVHHKTLGIVGMRIGMAL
 AQRVHGFNMPILYNARGHKBEARNARYCDLTLOESDFVCLLPLTDETHLUF
 GAQFAKMSAIFINARGGPVVDENALIAALOKGEIHAAGLDVPEGEPLSVDSFLLS
 MANVAVPHIGSAITHETRYGMAACAVNDLIDALQGVKKNVPHVAD"
 complement (3293..3919)
 /gene="yiaD"
 /locus_tag="S4183"
 complement (3293..3919)
 /gene="yiaD"
 /locus_tag="S4183"
 /function="putative membrane; Not classified"
 /note="residues 1 to 208 of 208 are 88.94 pct identical to
 residues 12 to 219 of 219 from Escherichia coli K-12 :
 B3552"
 /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein"
 /protein_id="AAP19151.1"
 /db_xref="GI:30043431"
 /translation="MSGALVSGCTTNPYTGEEAGKSAIGAGLSLVGAGIGALSSS
 KDRGKALITGAAGALGGVGYMDVQEAKLDRMRGTGVSVTRSGDNIILNPNP
 VTFDSSAPLTPAGANTLTGVAAMLKEYPKTAVNVITGYTDTGCHDLNMLSLQORADS
 VASALITQGVDSRIRITQGLGPANPIASNTABGKAQNRREITLSPL"
 4160..6439
 /gene="bisC"
 /locus_tag="S4184"
 4160..6439
 /gene="bisC"
 /locus_tag="S4184"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 Biotin"
 /note="residues 21 to 759 of 759 are 98.78 pct identical

gene

CDS

/product="hypothetical protein"
 /protein_id="AAP19150.1"
 /db_xref="GI:30043430"
 /translation="MPEVILYKALPDDLQRLQAHFTVHVQANLSPTVVQNAIPA
 EAELGSGNENVDALILEKPKLRATSTISGYDNFDVDALTKAKILMHTPTVLITET
 VADTLMALVLTARRVVEAERVKAGEWTASIGDWTGTDVHHKTLGIVGMRIGMAL
 AQRVHGFNMPILYNARGHKBEARNARYCDLTLOESDFVCLLPLTDETHLUF
 GAQFAKMSAIFINARGGPVVDENALIAALOKGEIHAAGLDVPEGEPLSVDSFLLS
 MANVAVPHIGSAITHETRYGMAACAVNDLIDALQGVKKNVPHVAD"
 complement (3293..3919)
 /gene="yiaD"
 /locus_tag="S4183"
 complement (3293..3919)
 /gene="yiaD"
 /locus_tag="S4183"
 /function="putative membrane; Not classified"
 /note="residues 1 to 208 of 208 are 88.94 pct identical to
 residues 12 to 219 of 219 from Escherichia coli K-12 :
 B3552"
 /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein"
 /protein_id="AAP19151.1"
 /db_xref="GI:30043431"
 /translation="MSGALVSGCTTNPYTGEEAGKSAIGAGLSLVGAGIGALSSS
 KDRGKALITGAAGALGGVGYMDVQEAKLDRMRGTGVSVTRSGDNIILNPNP
 VTFDSSAPLTPAGANTLTGVAAMLKEYPKTAVNVITGYTDTGCHDLNMLSLQORADS
 VASALITQGVDSRIRITQGLGPANPIASNTABGKAQNRREITLSPL"
 4160..6439
 /gene="bisC"
 /locus_tag="S4184"
 4160..6439
 /gene="bisC"
 /locus_tag="S4184"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 Biotin"
 /note="residues 21 to 759 of 759 are 98.78 pct identical

gene

CDS

/product="hypothetical protein"
 /protein_id="AAP19150.1"
 /db_xref="GI:30043430"
 /translation="MPEVILYKALPDDLQRLQAHFTVHVQANLSPTVVQNAIPA
 EAELGSGNENVDALILEKPKLRATSTISGYDNFDVDALTKAKILMHTPTVLITET
 VADTLMALVLTARRVVEAERVKAGEWTASIGDWTGTDVHHKTLGIVGMRIGMAL
 AQRVHGFNMPILYNARGHKBEARNARYCDLTLOESDFVCLLPLTDETHLUF
 GAQFAKMSAIFINARGGPVVDENALIAALOKGEIHAAGLDVPEGEPLSVDSFLLS
 MANVAVPHIGSAITHETRYGMAACAVNDLIDALQGVKKNVPHVAD"
 complement (3293..3919)
 /gene="yiaD"
 /locus_tag="S4183"
 complement (3293..3919)
 /gene="yiaD"
 /locus_tag="S4183"
 /function="putative membrane; Not classified"
 /note="residues 1 to 208 of 208 are 88.94 pct identical to
 residues 12 to 219 of 219 from Escherichia coli K-12 :
 B3552"
 /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein"
 /protein_id="AAP19151.1"
 /db_xref="GI:30043431"
 /translation="MSGALVSGCTTNPYTGEEAGKSAIGAGLSLVGAGIGALSSS
 KDRGKALITGAAGALGGVGYMDVQEAKLDRMRGTGVSVTRSGDNIILNPNP
 VTFDSSAPLTPAGANTLTGVAAMLKEYPKTAVNVITGYTDTGCHDLNMLSLQORADS
 VASALITQGVDSRIRITQGLGPANPIASNTABGKAQNRREITLSPL"
 4160..6439
 /gene="bisC"
 /locus_tag="S4184"
 4160..6439
 /gene="bisC"
 /locus_tag="S4184"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 Biotin"
 /note="residues 21 to 759 of 759 are 98.78 pct identical

to residues 1 to 739 from Escherichia coli K-12 :
B3551"
/codon_start=1
/transl_table=11
/product="biotin sulfoxide reductase"
/protein_id="AAP19152.1"
/db_xref="GI:30043432"
/translation="MLVETDGTFTVSSRGALATGWNLSQSAVRQVHNSHTRVPPMV
RKGFLASPENFGIRQDFVRSWDEALDLHQHQRIRREAYGPASIFAGSYGRSN
GVLRKATLLQKALAGYTGHELDYSTGAQAIIMPYVGVSGEYVQOQTWPLVLEH
SDVVWJANGANTLKIWNASDEQGLSYFSALRDSGKKLICIDPMRSETVDFGDKM
EWAVPMGTDMALMIGIAHTLVENGWHDDEAFIARCTTGVAFVASYLLGESDIKAAAE
WAAETCGCAAKRLAAIFHONTTLMAGWQWQFQGEQKHMIIVTLAAMLGQIGT
PGGGLSVHFANGNPTRSAVLSMKQSLPGGDAVDKIPARIVALEALNPGAYQ
HNGMRHPEDNFIWAGGANTHEDQINRLRARKWELVVISCEPWIAAKHADIY
LPATTSFPRDITDVSNGHLVPMQVPPRYEARNDQVFAELSRWKEGGYARF
TEGSQLQWLETFTYVAVRQASQVLEPPFAEFQWQNLMPENPSERPIRFADP
CSDPLQWLTAGSIEIFSRADYGPDCFGHMPLEPDEMQWASPEQLVLSAH
PAHLHLSQNLNYSLELAVANREPVIHPDQAQGITEGDMVRVWNSRGOILAGAV
ISEGKPGVICTHGAWPDLDUTADGICKNGAVNVLTKDLPSSRLGNCAGNTLWAL
EYNGPELTLTAPEPPAS"
complement(6408..6848)
/gene="yiac"
/locus_tag="S4185"
complement(6408..6848)
/gene="yiac"
/locus_tag="S4185"
/note="residues 1 to 146 of 146 are 97.94 pct identical to
residues 1 to 146 of 146 from Escherichia coli K-12 :
B3550"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP19153.1"
/db_xref="GI:30043433"
/translation="MIREAQSELPALILEMESTTGWHPFKYANWRDCIPLVRDAY
LANAQWEEQGLKGFVSMERFLAAMPVAPKAVRGIGKALMQYVQORYPHLM
EYVQKQPAIDPYRAGGHHVDCAWQDETQLEPTWINSWPVQTL"
complement(6845..7408)
/gene="tag"
/locus_tag="S4186"
complement(6845..7408)
/gene="tag"
/locus_tag="S4186"
/function="enzyme; DNA - replication, repair,
restriction/modification"
/note="residues 1 to 187 of 187 are 97.32 pct identical to
residues 1 to 187 of 187 from Escherichia coli K-12 :
B3549"
/codon_start=1
/transl_table=11
/product="3-methyl-adenine DNA glycosylase I"
/protein_id="AAP19154.1"
/db_xref="GI:30043434"
/translation="MERCQVWSQPLVIAYHDNMGVPEVTDKLFEMICFEGQOAGL
SWITVLKRENYRAYFHQDPVKVAAQOEDEVERLVQDAGLIIRHGKIOALIGNRAY
IQMEQNGEPFPVPSVFNHQPVQVQTQTLSEIPTSTASALSALKAKRGKFKVGT
ICISFMQACGLVNDHVVGCCYLGKRP"
7566..8264
/gene="yhjY"
/locus_tag="S4187"
7566..8264
/gene="yhjY"
/locus_tag="S4187"
/function="putative enzyme; Not classified"
/note="residues 1 to 232 of 232 are 99.13 pct identical to
residues 1 to 232 of 232 from Escherichia coli K-12 :
B3548"
/codon_start=1
/transl_table=11
/product="glutathione oxidoreductase"
/protein_id="AAG58632.1"
/db_xref="GI:12518197"
/translation="MTKHYDYIAIGGGSGGIASINRAAMYCQKALIEAKELGGTCVN
LGVKNVDVIKGFARFVDKATLEVNGTITADHILJATGSRPSHEDIPIGVYGDSDGP
FALPALPFRVAVVAGYLAELVAGVINGLCAKTHLFVRKHAPLSPDPMISRTVEVM
VAKNKGITVVDKIQNTNVEGTVAVGNTGAVETVPVAAAGRLSERLFPNNKPEHL
DYGNIPFVSPHPPIGTVGLTEPQARQYGDQVKYKYSFTAMYTAVTTTRQPCRMK
XVCVGPBEKIVGIGIGFGNDEMQLQGFALVQKATKDFDNTVAIHPTAAEEFVTR
2300..2653
/gene="arsR"
/note="synonym: Z4903"
2300..2653
/gene="arsR"

Query Match

Best Local Similarity

Matches 42; Conservative

100.0%; Score 42; DB 1; Length 289816;

Pred. No. 0.00025; Mismatches 0; Indels 0; Gaps 0;

1 CTGCACCTACACATTCGTTAAGTCATATATGTTTGTGACTTA 42

100.0%; Score 42; DB 1; Length 289816;

Pred. No. 0.00025; Mismatches 0; Indels 0; Gaps 0;

1 CTGCACCTACACATTCGTTAAGTCATATATGTTTGTGACTTA 42

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

```

/function="regulator; Protection responses: Drug/analog
sensitivity"
/notes="Residues 1 to 117 of 117 are 94.87 pct identical to
residues 1 to 117 of 117 from Escherichia coli K-12 Strain
MG1655: B3501"
/codon_start=1
/transl_table=11
/product="transcriptional repressor of chromosomal ars
operon"
/protein_id="AAG58633.1"
/db_xref="GI:12518198"
/translation="MSFLPIQLFKILADETRIGVILLSELGELVCVCDLCTALDQSQ
PKISRLALRESGLLDRKQKQWVHRLSPHPSWAALIEQAWRCEKEKQVIVRN
LAKQNCVDSKNTCS"
2707..3996
/genes="arsB"
/notes="synonym: Z4904"
2707..3996
/genes="arsB"
/function="transport; Protection responses: Drug/analog
sensitivity"
/notes="Residues 1 to 429 of 429 are 100.00 pct identical
to residues 8 to 436 of 436 from Escherichia coli K-12
Strain MG1655: B3502"
/codon_start=1
/transl_table=11
/product="arsenical pump membrane protein"
/protein_id="AAG58634.1"
/db_xref="GI:12518199"
/translation="MLLAGAIFVLTIVLWOPKGGIGKWSATGLAVLALVGVVHPG
DIPVWNIWNNATAAPIAVIIISLLDESPEFMAALHVSRWNGRGLLPTWVLLG
AAVAFNGDGAALITPIVIAMLLALGFSKGTTLAFVMAAGFIADTASLPTIVNLV
NIVSADFGLGREYASVAVPVDIAAIVATLVMLHYFRKIPONDYDMLAKSPARAI
KQPAETFGWVLLGVLFVLEPIGIPVSAIAAVGALILFVAKRGHAIATGKVLIR
GAPQWIVISGLMYLVYGLRNAGLLEYLSGVNLVADNGLWAATLCTGFLPAFLSSI
KNNMPTVLGALSIDGASGVKEAWVTANVIGCDLGPKEITPIGSLATLLWLHLVLSQ
KNTISWGYFFRTIITLPVLPVTLAALALRSFTL"
4009..4434
/genes="arsC"
/notes="synonym: Z4905"
4009..4434
/genes="arsC"
/function="enzyme; Protection responses: Drug/analog
sensitivity"
/notes="Residues 1 to 141 of 141 are 98.58 pct identical to
residues 1 to 141 of 141 from Escherichia coli K-12 Strain
MG1655: B3503"
/codon_start=1
/transl_table=11
/product="arsenate reductase"
/protein_id="AAG58635.1"
/db_xref="GI:12518200"
/translation="MSNITVHNPAQGTSRNTELMIRNSGTPTIHYLETPTRDEL
VKLIADMGISVRLAKNVPVEELGLADKFTDDRLLIDFMLQHPILINRPIVWTLG
TRLCRSEVVLEILPDAQGATKEGKVDENGRLK"
4546..4653
/genes="Z4906"
4546..4653
/genes="Z4906"
/function="orf; Unknown function"
/notes="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG58636.1"
/db_xref="GI:12518201"
/translation="MLFVLGTGVAHQAILFTHYSUREIPITASSMMYQ"
5062..6177
/genes="Z4907"
5062..6177
/genes="Z4907"
/function="orf; Unknown function"
/notes="Residues 1 to 249 of 371 are 89.95 pct identical to
residues 1 to 249 of 260 from Escherichia coli K-12 Strain

```

```

MG1655: B3504"
/codon_start=1
/transl_table=11
/protein_id="AAG58637.1"
/db_xref="GI:12518202"
/translation="MSIDFTFGMINTYHGIYNRITDTENVKTPDTPMWPCDNREBQO
FINSTGEGINPEQIDARQHLPOIYACRINKYTDADYSYVQSLVSLITNISIS
STPLRQAEIILNQINNIYKKNKSAECRIIVANPKPKDAIITKISVEEGIPVRFVS
TQNFSDTFPIABQADLFTNIKDQSLYQKMTGLYIEHSENKRMKVAGTGFIDFNM
TQNLGSGFVLTLSPFEDLLNINFTDANLGDITFSHKEHPTPKLYKQBYLDKQTEG
LPSTLLTNDNPLRAKASTIATIKPLBARITNLSYXDLILKYQBFQKCYQVKSFT
TLRSYVKIQTWAEENSEYQFEVQYETLNPCKCHARLI"
6499..7098
/genes="elp"
/notes="synonym: Z4908"
6499..7098
/genes="elp"
/function="membrane; Cell envelop: Outer membrane
constituents"
/notes="Residues 1 to 199 of 199 are 99.49 pct identical to
residues 1 to 199 of 199 from Escherichia coli K-12 Strain
MG1655: B3506"
/codon_start=1
/transl_table=11
/product="outer membrane protein induced after carbon
starvation"
/protein_id="AAG58638.1"
/db_xref="GI:12518203"
/translation="WVLIFVDRDSNMNTKALLISLFLAACSISIPONIKGNQPD
IQKSVAVHNPGLYVQQOARPGGVINVGKTDLTLEIAVLPLDSYAKPDIEANYQ
GRLRLAGSGFLDPVNYRNVHTILGTIOGQPGFINKVPYFNFLYMNQGIQVHILREV
VNTTYLMDYGYGAFWPGWPGWAPYTTNAVSQVTELVK"
7254..7784
/genes="yhiP"
/notes="synonym: Z4909"
7254..7784
/genes="yhiP"
/function="orf; Unknown function"
/notes="Residues 1 to 176 of 176 are 99.43 pct identical to
residues 1 to 176 of 176 from Escherichia coli K-12 Strain
MG1655: B3507"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG58639.1"
/db_xref="GI:12518204"
/translation="WELIITRDTMPTAMKNIKSGNVVHIONEIEIDVMLHONAFVI
IDTLNMYFHSNFLTQIEELKPVHVIIEPNIKEKLGKVPVTFYRTITIDFVALI
KGSYCVPEANVSLSRKHQVLSCTIANOMTIEDILEKLSIKTYCHKHNMIMNL
KRINELVRHQHIDYLV"
7786..>11071
/notes="O-island #140; Region of the BDL933 chromosome not
homologous to E. coli K-12 MG1655"
/complement(7845..8873)
/genes="chus"
/notes="synonym: Z4910"
/complement(7845..8873)
/genes="chus"
/function="putative transport; Transport of small
Query Match 88.6%; Score 37.2; DB 1; Length 11071;
Best Local Similarity 92.9%; Pred. No. 0.016;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 CTGCACCTTACACATCGGTAACTATATATATGTTTGGACTTA 42
|||||
DB 2213 CTGCACATACACATCGGTAAACCATATATGTTTGGACTTA 2254
|||||

```

```

RESULT 9
AP002565 257988 bp DNA linear BCT 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 16/20.
DEFINITION

```

```

ACCESSION      AP002565 BA000007
VERSION        AP002565.1 GI:13363693
KEYWORDS       Escherichia coli O157:H7
SOURCE         Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
ORGANISM       Enterobacteriaceae; Escherichia.
REFERENCE      1 (sites)
AUTHORS        Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
                Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
                Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
                Sasakawa,C. and Shinagawa,H.
TITLE          Complete nucleotide sequence of the prophage VT2-Sakai carrying the
                verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
                derived from the Sakai outbreak
JOURNAL        Genes Genet. Syst. 74 (5), 227-239 (1999)
MEDLINE        20198780
PUBMED         10734605
REFERENCE      2 (sites)
AUTHORS        Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
                Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
                Hayashi,T.
TITLE          Comparative analysis of the whole set of rRNA operons between an
                enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
                Escherichia coli K-12 strain MG1655
JOURNAL        Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
MEDLINE        20557356
PUBMED         11198008
REFERENCE      3 (sites)
AUTHORS        Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
                Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
                Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
                Shinagawa,H.
TITLE          Complete nucleotide sequence of the prophage VT1-Sakai carrying the
                Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
                O157:H7 strain derived from the Sakai outbreak
JOURNAL        Gene 258 (1-2), 127-139 (2000)
MEDLINE        20564182
PUBMED         11111050
REFERENCE      4 (sites)
AUTHORS        Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
                Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
                Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
                Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
                Shinagawa,H.
TITLE          Complete genome sequence of enterohemorrhagic Escherichia coli
                O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL        DNA Res. 8 (1), 11-22 (2001)
MEDLINE        21156231
PUBMED         11258796
REFERENCE      5 (bases 1 to 267888)
AUTHORS        Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
                Hayashi,T.
TITLE          Direct Submission
JOURNAL        Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
                Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
                Japan (E-mail:kenegen-info.osaka-u.ac.jp,
                URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
                Fax:81-6-6879-2047)
COMMENT        genome project.
FEATURES       Location/Qualifiers
                1..267888
                /organism="Escherichia coli O157:H7"
                /mol_type="genomic DNA"
                /strain="O157:H7"
                /sub_strain="RIMD 0509952"
                /db_xref="taxon:83334"
                151..318
                /gene="ECs4220"
                151..318
                /gene="ECs4220"
                /note="Identical to YHFL_ECOLI gi|1789770 (Conserved in
                E.coli K-12)"

gene
CDS

/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37643.1"
/db_xref="GI:13363694"
/translation="MNRKIFKVALGVAVLATLTACTGHIENRDRKNCSDYLLHPLAIS
KIIGCGPTAQ"
563..1951
/gene="ECs4221"
563..1951
/gene="ECs4221"
/note="similar to YHFM_ECOLI gi|2367216 percent identity
99 in 462 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative amino acid/amine transport protein"
/protein_id="BAB37644.1"
/db_xref="GI:13363695"
/translation="MTNSPLQVQGEKGIAMSGQELQRLGFWAVLAIAGVTTVGSG
IFVSGVAKAAGTFLTTLVLAFLVIGLIVIPQCVYAEALSTAYPENAGDYVYLKNGS
RPLAFLGWFASFWANDAPSLISALAIIVSNLGLTIDPLIGKFAAGLIAPMLLHL
RSVGGAAFTLIATIAIIPFTIVIGLIFWKAENPAAPATTAIGATGSPMALLAGI
SATSMYTGWASICYWTGRIKNGKTPRALIGSLCLLVLTLYLALVLSGLMPDKL
ANSETPISDALTHIIPALGSTAGIFVAITAMIVILGSLSSCVMYQPRLEYAKAKNEFF
KCPGHVPKNTPDVSILOALGIFPIFVSDUTSLIGIFTLVMCFKNTLTPGSLIMC
RRDDYKPLMRTPAFLMTPLATASSLILVASTFVWAPIPGLICAVIVATGFLPAYAF
WAKESRQLNALS"
1951..2994
/gene="ECs4222"
1951..2994
/gene="ECs4222"
/note="similar to YHFN_ECOLI gi|1789772 percent identity
98 in 347 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative amino acid/amine transport protein"
/protein_id="BAB37645.1"
/db_xref="GI:13363696"
/translation="MJSGENKMLDIDKSTVDLVTENMQVEKVLSHDVLVHTTIVE
EMVKRIDRIYFVACGSLNAAQTAHLADRFSDLOVYASGHEFCNTPYRLDARCA
VIGSDYKTEVTEIKALELGRACGALTAAFTKRAADSPITSAASFSDIYQADCIWEIHL
LLCVSVLEMTIRLAPHAEGIKNDLKPALGNALGVLTVRWEGKRGGLASOWPMI
VYAAAGPLRPLYGKGIIVTMEFTWTHGCVIESGEPHGLPVEIVGCVPLFLGNDG
SRITERAINFVKQRTDNVIDVYALSQSLHPLAPFLMFPVMEHLCYLSIYKDN
PDERRIYGLVSEI"
3045..3872
/gene="ECs4223"
3045..3872
/gene="ECs4223"
/note="N-terminal part (1-115 in 294 aa) is similar to
YHFO_ECOLI gi|1789773 percent identity 97 in 115 aa,
C-terminal part (162-294 in 294 aa) is similar to
YHFP_ECOLI gi|2367217 percent identity 99 in 133 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37646.1"
/db_xref="GI:13363697"
/translation="MKTGMTGQHQRLPIEHAFRDASSELGYDGIETWGRPHAPAPDL
KAGGIKQIKALAQYQMPITIGYPTNGYPNNMLGDEHRRRESLDMIKLAWAKEM
NAGYTLISAHAGYLTTPNNVWGRLAENSELCEYANIGMDLILEPLTYPNSVGN
ANDVLHALALVPSPRLPSVDVICAPYVQAFPMVSPFDKLGDLHLHLIVSDGASDTH
YIPGEGKPLRELMRDIIDRGYEGYCVSELVITMYNEPRLYARQALRFRALLPEDE"
3872..4204
/gene="ECs4224"
3872..4204
/gene="ECs4224"
/note="similar to N-terminal part (1-110 in 261 aa) of
YH5Q_ECOLI gi|1789775 percent identity 98 in 110 aa

```



```
* 119393 186963: contig of 67571 bp in length
* 186964 187063: gap of 100 bp
* 187064 198631: contig of 11568 bp in length.

FEATURES
    Location/Qualifiers
        1..198631
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /clone="RP23-29119"
            /clone_lib="RPCI-23 Female Mouse BAC"
        misc_feature 1..86971
            /note="assembly_fragment"
            /clone_end:SP6
            vector side:left"
        misc_feature 87072..88053
            /note="assembly_fragment"
        misc_feature 88154..90343
            /note="assembly_fragment"
        misc_feature 90444..98277
            /note="assembly_fragment"
        misc_feature 98378..105014
            /note="assembly_fragment"
        misc_feature 105115..119292
            /note="assembly_fragment"
        misc_feature 119393..186963
            /note="assembly_fragment"
        misc_feature 187064..198631
            /note="assembly_fragment"
            /clone_end:T7
            vector side:right"
```

ORIGIN

```
Query Match      61.0%; Score 25.6; DB 2; Length 198631;
Best Local Similarity 77.5%; Pred. No. 83;
Matches 31; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
```

Qy
Db

1 CTGCACCTTACACATTTCGTTAAGTCATATATGTTTTTGACT 40
|||||
285730 CTC CACTTATACATTGTGTTTATACACATATATATTGACT 185691

[illegible]

REFERENCES

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 88013)

Rouken, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dore, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D. and Hood, L.

Sanger, F. and Coulson, A.R. 1975. Sequencing of human chromosome 15 D1S5146-D15S117 region

TITLE	Sequencing of human chromosome 15 DISC148-D15S117 region
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 88013)
AUTHORS	Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A., Nesbitt,R., Traicoff,R. and Hood,L.
TITLE	Direct Submission
JOURNAL	Submitted (27-FEB-2001) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA
REFERENCE	3 (bases 1 to 88013)
AUTHORS	Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R., Pate,D. and Hood,L.
TITLE	Direct Submission
JOURNAL	Submitted (14-JUN-2001) Multimegabase Sequencing Center, Institute

REFERENCE	AUTHORS	TITLE	JOURNAL
3	(bases 1 to 150965)	DOE Joint Genome Institute and Stanford Human Genome Center. Direct Submission	
	Submitted (29-NOV-2001)	DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
4	(bases 1 to 150965)	DOE Joint Genome Institute and Stanford Human Genome Center. Direct Submission	
	Submitted (21-DEC-2001)	DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
	On Dec 21, 2001 this sequence version replaced gi:17149379. Draft Sequence Produced by DOE Joint Genome Institute www.jgi.doe.gov		
REFERENCE	AUTHORS	TITLE	JOURNAL
COMMENT			

Note: Data from AC010999 [Drafting center: UWMSC], AC079076 [Drafting center: WIBR], AC016554 [Drafting center: UWMSC] and AC090532 [Drafting center: WIBR and UWMSC] were added for finishing

```

-----
FEATURES
source
Location/Qualifiers
1. .88013
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="15"
   /map="15q21.3"
   /clone="CTD-237008"
   /clone_lib="Cal Tech human BAC library D"
   /notes="This clone overlaps RP11-323F24 AC010999,
RP11-347P18 AC079076, RP11-559N7 AC016554 and RP11-325L11
AC090532. Data from these BACs were added and the
consensus sequence determined from CTD-237008 to the
extent possible"
misc_feature
1. .56308
   /note="overlap with RP11-559N7 AC016554"
misc_feature
1. .25594
   /note="overlap with RP11-323F24 AC010999"
misc_feature
21346. .88013
   /note="overlap with RP11-347P18 AC079076"
unsure
65975. 65985
   /note="low quality data"
misc_feature
73351. .88013
   /note="overlap with RP11-325L11 AC090532"

```

```

Query Match.      59.5%; Score 25; DB 9; Length 88013;
Best Local Similarity 75.6%; Pred. No. 1.5e+02;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy      2      TGCACCTTACACATTGGTTAAGTCATATATGTTTTGACTTA 42
          ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      36266  TGCACCTTACACATTGGTTATGATGTTGATGTTTTGCTTAA 36306
          ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 14	
LOCUS	AC091915 150965 bp DNA linear PRI 21-DEC-2001
DEFINITION	Homo sapiens chromosome 5 clone RP11-224A9, complete sequence.
ACCESSION	AC091915
VERSION	AC091915.4 GI:17975242
KEYWORDS	HTG.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 150965)
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 150965)
AUTHORS	DOE Joint Genome Institute.
TITLE	Direct Submission
JOURNAL	Submitted (09-JUN-2001); Production Sequencing Facility, DOE Joint

JOURNAL
 Direct Submission
 Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (13-FEB-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 169377)
 Birren,B., Nuebaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
 Barina,N., Bastien,V., Bloom,F., Boguslavskiy,L., Boukhalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
 Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Fero,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
 Hago,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
 Matthews,C., McCarthy,M., Meldrim,J., Menus,L., Minova,I.,
 Mlepta,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
 Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
 Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R.,
 Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Tallamas,J., Tesfaye,S., Theodore,J., Topham,K.,
 Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (19-OCT-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Oct 19, 2002 this sequence version replaced gi:20455668.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: l23101
 Center clone name: l11_G3
 ----- Summary Statistics
 Sequencing vector: Plasmid; N/A; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 169119 bases at least Q40
 Consensus quality: 169139 bases at least Q30
 Consensus quality: 169232 bases at least Q20
 Insert size: 168000; agarose-fp
 Insert size: 169277; sum-of-contigs
 Quality coverage: 18.3 in Q20 bases; agarose-fp
 Quality coverage: 18.1 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submittor.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 157396: contig of 157396 bp in length
 * 157397 157496: gap of 100 bp
 * 157497 169377: contig of 11881 bp in length.
 Location/Qualifiers
 1..169377
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clone="RP11-111G3"
 /clone_lib="RP11-11 Human Male BAC"
 i. .157396
 misc_feature
 /note="assembly_fragment"

clone_end:SP6
 vector_side:left
 157497..169377
 /note="assembly_fragment"

 misc_feature
 ORIGIN
 Query Match 59.5%; Score 25; DB 2; Length 169377;
 Best Local Similarity 75.6%; Pred. No. 1.4e+02;
 Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 Oy 2 TGCACCTTACACATTCGTTAAGTCATATATGTTTTTGGACTTAA 42
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 60812 TGCACCTTACACATTCGTTAAGTCATATATGTTTTTGGACTTAA 60852
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: May 26, 2004, 17:22:12
 Job time : 591.231 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 15:20:44 ; Search time 556.41 Seconds
(without alignments)
3115.905 Million cell updates/sec

Title: US-10-676-299-6

Perfect score: 40

Sequence: 1 taagtcaaaacatatatgacttaacgaatgtgaagtgc 40

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	40	100.0	3492	1	ECARSRC	X80057 E.coli gene
C 2	40	100.0	10240	1	AE000426	AE000426 Escherich
C 3	40	100.0	11524	1	AE015361	AE015361 Shigella
C 4	40	100.0	179941	2	AC145934	AC145934 Gallus chr
C 5	40	100.0	225419	1	EC00W76	U00039 E. coli chr
C 6	40	100.0	242495	2	AC146183	AC146183 Pan trogi
C 7	40	100.0	289816	1	AE016992	AE016992 Shigella
C 8	35.2	88.0	11071	1	AE005575	AE005575 Escherich
C 9	35.2	88.0	267888	1	AP002565	AP002565 Escherich
C 10	31.6	79.0	301660	1	AE016768	AE016768 Escherich
C 11	24.6	61.5	166112	2	AC102260	AC102260 Mus muscu
C 12	24.6	61.5	176496	10	AL772194	AL772194 Mouse DNA
C 13	24.2	60.5	90650	10	AF242431	AF242432 Mus muscu
C 14	24.2	60.5	156569	2	AC110190	AC110190 Homo sapi
C 15	24.2	60.5	164314	9	AC091691	AC091691 Homo sapi
C 16	24.2	60.5	194366	5	AL954179	AL954179 Zebrafish
C 17	24.2	60.5	198631	2	AC116741	AC116741 Mus muscu
C 18	24	60.0	88013	9	AC090511	AC090511 Homo sapi
C 19	24	60.0	150965	9	AC091915	AC091915 Homo sapi
C 20	24	60.0	169377	2	AC110578	AC110578 Homo sapi
C 21	24	60.0	178127	2	AC079076	AC079076 Homo sapi
C 22	24	60.0	198719	2	AC015716	AC015716 Homo sapi
C 23	24	60.0	207471	10	AC122249	AC122249 Mus muscu
C 24	24	60.0	212691	2	AC118594	AC118594 Mus muscu
C 25	23.8	59.5	214795	2	AC110449	AC110449 Rattus no
C 26	23.8	59.5	233330	2	AC112548	AC112548 Rattus no
C 27	23.6	59.0	87548	5	BX323038	BX323038 Zebrafish
C 28	23.6	59.0	101904	5	AL935310	AL935310 Zebrafish
C 29	23.6	59.0	122100	8	AP005774	AP005774 Oryza sat
C 30	23.6	59.0	135070	2	AP003741	AP003741 Oryza sat
C 31	23.6	59.0	142711	10	AC121863	AC121863 Mus muscu
C 32	23.6	59.0	149928	2	AC134950	AC134950 Danio rer
C 33	23.6	59.0	152433	2	AP004297	AP004297 Oryza sat
C 34	23.6	59.0	164936	5	BX255954	BX255954 Zebrafish
C 35	23.6	59.0	168990	2	AC122516	AC122516 Mus muscu
C 36	23.6	59.0	183417	5	AL845320	AL845320 Zebrafish
C 37	23.6	59.0	184808	2	AC118619	AC118619 Mus muscu
C 38	23.6	59.0	186668	2	AC114627	AC114627 Mus muscu
C 39	23.6	59.0	196688	2	BX323087	BX323087 Danio rer
C 40	23.6	59.0	209383	2	AC133953	AC133953 Mus muscu
C 41	23.6	59.0	221146	10	AC098739	AC098739 Mus muscu
C 42	23.6	59.0	224806	2	BX004858	BX004858 Danio rer
C 43	23.6	59.0	225082	10	BX548065	BX548065 Mouse DNA
C 44	23.6	59.0	226871	2	AC098604	AC098604 Rattus no
C 45	23.6	59.0	253066	2	AL935184	AL935184 Danio rer

ALIGNMENTS

RESULT 1	ECARSRC	3492 bp	DNA	linear	BCT 20-JUL-1995
ECARSRC/c	ECARSRC				
LOCUS	E.coli genes arsr, arsb, arsc.				
DEFINITION	X80057				
ACCESSION	X80057.1	GI:510824			
VERSION	arsB gene; arsc gene; arsenate reductase; arsenic-efflux pump;				
KEYWORDS	arsenic-inducible repressor; arsr gene.				
SOURCE	Escherichia coli				
ORGANISM	Escherichia coli				
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
AUTHORS	Enterobacteriaceae; Escherichia.				
TITLE	1				
	Diorio,C., Cai,J., Marmor,J., Shinder,R. and DuBow,M.S.				
	An Escherichia coli chromosomal ars operon homolog is functional in				

```

/transl_table=11
/product='arsenate_reductase'
/protein_id='CMA56363.1'
/db_xref='GI:516212'
/db_xref='GOA:P37311'
/db_xref='SWISS-PROT:P37311'
/translation='MSNITVYNPACGTSGNTLEMRNSGTETPTIIHYLETPTPEDEL
VKLIADMGISVRALLKRNKVEPYEELGAEKPTDDELIDFMQLQHPILINRPVIVTPLG
TLCRPSVVLEILPDAQGAFSGKEDGKXVDEAGKRLK"
repeat_region 2892..2905
/rpt_type=INVERTED
repeat_region 2910..2923
/rpt_type=INVERTED
ORIGIN
Query Match 100.0%; Score 40; DB 1; Length 3492;
Best Local Similarity 100.0%; Pred. No. 0.0022;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TAAGTCAAAACATATATGACTTAAACGAATGCTAAAGTGC 40
|||||
Db 662 TAAGTCAAAACATATATGACTTAAACGAATGCTAAAGTGC 623

RESULT 2
AE000426/c
LOCUS AE000426 10240 bp DNA linear BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 316 of 400 of the complete
genome.
ACCESSION AE000426 U00096
VERSION AE000426.1 GI:11789910
KEYWORDS .
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1 (bases 1 to 10240)
Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
Mau,B. and Shao,Y.
The complete genome sequence of Escherichia coli K-12.
Science 277 (5331), 1453-1474 (1997)
97424617
9278503
2 (bases 1 to 10240)
Blattner,F.R.
Direct Submission
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
3 (bases 1 to 10240)
Blattner,F.R.
Direct Submission
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
4 (bases 1 to 10240)
Plunkett,G. III.
Direct Submission
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using GeneMark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG

```

Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES	source	Location/Qualifiers
promoter	gene	1. .10240 /organism="Escherichia coli K12" /mol_type="genomic DNA" /strain="K12" /sub_strain="WGI655" /db_xref="taxon:83333" <i>.31 /note="factor Sigma70; predicted +1 start at 3638369"
		161. .1630 /gene="yhiP" /note="synonym: b3496" 161. .1630 /gene="yhiP"
		/function="putative transport; Not classified" /note="o489; 99 pct identical amino acid sequence and equal length to YHIP_ECOLI SW: P36837" /codon_start=1 /transl_table=11 /product="putative transport protein" /protein_id="AAC76521.1" /db_xref="GI:1789911"
CDS	gene	/translation="MNTTTPMGLQPPFFMIFFVELWRFYGVQGVLAFFVFKQ LGFQSOQAFVFGAALVYGLISIGVGVGDLHLLTKETIVLGLALVLAIGVFMQMSL LKPDILFIALGTIAVGNGLFKANPASLLSKVPPKPRLDGAPTLFYMSINIGSLAL SLAPVADRQGVTVNYLGGAGLIIALLAVIACRCMWKDIGSBPDPRMSPSKLLYL LGSVMIIFVCALMHNVEANLVILVLSIVTVIIFFRQAFKLDKTRKMFVAVML EAVFYILYQPTSLNFAINNVHHEILGFSINPVSPQALNPFVAVLASFILAGIYT HLGNKDLSPMPKFTLGMFNCSLGFLTAAGAKMFWADAQGLTFPFIIVLYLFSQSG BLFISALGLAMIALVPOHLMGFLGMWFLTQAAFLGGYVATFTAVPDNITDPLET LPVTVNFGKGLVTLGVAVVNLMLVPMWKMIATPESH" complement (1679. .2431) /gene="yhiQ" /note="synonym: b3497" complement (1679. .2431) /gene="yhiQ"
		/function="orf; Unknown" /note="f250; 100 pct identical to 218 amino acids of YHIQ_ECOLI SW: P37633 but has 32 additional N-ter residues; stg start, alternate starts possible" /codon_start=1 /transl_table=11 /product="orf, hypothetical protein" /protein_id="AAC76522.1" /db_xref="GI:1789912"
		/translation="WKICLIDETGTGDCALSVLAARMGLEHDEDNMALVLTPEHLEL RKBPDLGGIFVDFVGGMAHRRKFGGGEARVAKAVGIKGDVLPDVVDTAGLGRD AFVLASGCRVRLMERNFVVAALLDGLARGYADATGCGWLOERLQLIHASSLTALTD ITPRQVYLFDMFPKQKSAVKKRVFQSLVGPDLDDGLLEPARKLAKRVVVK RPDYAPPLIYATFNAAVTVKGRFDIYAGTPV" complement (2439. .4481) /gene="prlC" /note="synonym: b3498" complement (2439. .4481) /gene="prlC"
promoter	gene	/EC_number="3.4.24.70" /function="enzyme; Degradation of proteins, peptides, glyco" /note="f680; 100 pct identical to OPDA_ECOLI SW: P27298; CG Site No. 18031" /codon_start=1 /transl_table=11 /product="oligopeptidase A" /protein_id="AAC76523.1" /db_xref="GI:1789913"
		/translation="MTNPLTTPPELPFPFKILPEHVVPVAVTKALNDCRENVVVAQG APYTWENLCQFLAEVDVGLRIFSPVSHLSVKNPSPELREAYEQTLPLLSYSTWVGQ HEGLYKAYRDLRDGDHYATLNTAKKAVDNALRDFELSGICLPKQKQRYGEIATRLS SIGNOYNNVLDTMGWTKLVTDEALAGMPESAALAAQAARAKSEGLVLLTIDIPS YLPVMTYCDNALREEMVRAVSTRASDQGNAGWDSKVMEBILALRHELALQILGPE NYAFKSLSATMAENPQQVLDPLDTLAKRSPQCEKELAQARAFAPKRGVDELQPMDI AYTSKQKHYSISDEQLRPPYPPENKAVNKLFEVVRVIGITAKERKDVVWHPDVR FFELYDNBLRGSFYLDLYARKNKGAWMDCVQGMRRKADSGSLQKPFAYLTICNFR PVNGKPALFTHDEVITLTFHFHGLHMLTRITAGVSGISGVFWDAVEPLSPQFENW CWPEALAFISGHYETGEPLKSLLDMLAKNYQAAFLTRQLEGLFPFLHAEFR PDQGALETLETKLVAVVPSPWGRPPHAFSHIPAGGYAGGYLYLVADVLAADA FSRFEERGIINRETGQSFLDNLILSRGGSBEMDLFKFRGRREPQLDAMLEHYGIKG" complement (2457. .2488) /gene="prlC" /note="factor Sigma70; predicted +1 start at 3640781" complement (2545. .2575) /gene="prlC" /note="factor Sigma70; predicted +1 start at 3640869" complement (2580. .2608) /gene="prlC"
		/note="factor Sigma70; predicted +1 start at 3640904" complement (4546. .4573) /note="factor Sigma32; predicted +1 start at 3642870" 4561. .4587 /note="factor Sigma70; predicted +1 start at 3642925" 4684. .5526 /gene="yhiR" /note="synonym: b3499" 4684. .5526 /gene="yhiR"
CDS	gene	/function="orf; Unknown" /note="o280; 100 pct identical amino acid sequence and equal length to YHIR_ECOLI SW: P37634" /codon_start=1 /transl_table=11 /product="orf, hypothetical protein" /protein_id="AAC76524.1" /db_xref="GI:1789914"
		/translation="MLSYRHSFHAGNHADVLKHTVQSLIIIESKEKPKFLYLDTHAG AGRYOLGSEHAERTGEYLEGIAIWOODLPAELEAVINVYKPNRSGQLRYPGSGPL IASLLREQDSLQLTLPDSEFQKDSRAREKADGFOQLKAKLPVPVSRRL ILIDPPYEMKTDYQAVVSGIAGYKRPATGIYALWYPLVLRQQLKRIHLEATGIRK ILQLELAVLPDSDRRGMTASGMIVINPFWKLEOQMNVLFWLHSLKLVPAGTGHTVSW IVPE" 5291. .5303 /gene="yhiR"
		/note="central position to predicted promoter: -157.5" /bound moiety="Ada predicted site" 5421. .5449 /gene="yhiR" /note="factor Sigma70; predicted +1 start at 3643787" 5598. .6950 /gene="gor" /note="synonym: b3500" 5598. .6950 /gene="gor"
CDS	gene	/EC_number="1.6.4.2" /function="enzyme; Biosynthesis of cofactors, carriers: Thioedoxin, glutaredoxin, glutathione" /note="o450; 100 pct identical to GSHR_ECOLI SW: P06715; CG Site No. 665" /codon_start=1 /transl_table=11
protein_bind	promoter	

/product="glutathione oxidoreductase"
 /protein_id="AAC76525.1"
 /db_xref="GI:1789915"
 /translation="MTKHVDYIAIGSSGGIASINRAAYGKCALIBAKELGGTCVN
 VCGVKKVWMAAQI:REAHYMGPDYGTDTINKENWETLIASRTAYIDRLHTSVENY
 LCKNNVDVTKGPARVDKATLEVNGSTITADHILATGSRSHPDIPGVGIDSDGP
 PALPALPERVAVGAGYIAVELAGVINGLGAKTHLFPVKHAPLRSFDPMISSETLVEM

Query Match 100.0%; Score 40; DB 1; Length 10240;
 Best Local Similarity 100.0%; Pred. No. 0.0018;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TAAGTCAAAAACATATATGACTTAACGAATGTGTAAGTGC 40
 |||||
 Db 7781 TAAGTCAAAAACATATGACTTAACGAATGTGTAAGTGC 7742

RESULT 3
 AEO15361/c
 LOCUS
 DEFINITION Shigella flexneri 2a str. 301 linear BCT 18-OCT-2002
 genome.
 ACCESSION AEO15361 AB005674
 VERSION AEO15361.1 GI:24053992
 KEYWORDS
 SOURCE
 ORGANISM
 Shigella flexneri 2a str. 301
 Shigella flexneri 2a str. 301
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Shigella.
 1 (bases 1 to 11524)

Jin, Q., Yuan, Z.H., Xu, J.G., Wang, Y., Shen, Y., Lu, W.C., Wang, J.H.,
 Liu, H., Yang, J., Yang, F., Qu, D., Zhang, X.B., Zhang, J.Y., Yang, G.W.,
 Wu, H.T., Dong, J., Sun, L.L., Xue, Y., Zhao, A.L., Gao, Y.S., Zhu, J.P.,
 Kan, B., Chen, S.X., Yao, Z.J., He, B.K., Chen, R.S., Ma, D.L.,
 Qiang, B.Q., Wen, Y.M., Hou, Y.D. and Yu, J.

Genome sequence of Shigella flexneri 2a: insights into
 pathogenicity through comparison with genomes of Escherichia coli
 K12 and O157

Nucleic Acids Res. 30 (20), 4432-4441 (2002)
 12384590
 2 (bases 1 to 11524)
 Jin, Q., Shen, Y., Wang, J.H., Liu, H., Yang, J., Yang, F., Zhang, X.B.,
 Zhang, J.Y., Yang, G.W., Wu, H.T., Dong, J., Sun, L.L., Xue, Y.,
 Zhao, A.L., Gao, Y.S., Zhu, J.P., Chen, S.X., Yao, Z.J., Wang, Y.,
 Lu, W.C., Qiang, B.Q., Wen, Y.M. and Hou, Y.D.

Direct Submission
 Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry
 of Public Health, 100 Yingxin Jie, Xuanwu Qu, Beijing 100052, P.R.
 China

FEATURES
 Source
 1..11524
 /organism="Shigella flexneri 2a str. 301"
 /mol_type="genomic DNA"
 /strain="301"
 /serotype="2a"
 /db_xref="taxon:198214"
 67..1419
 /gene="gor"
 /locus_tag="SF3531"
 67..1419
 /gene="gor"
 /locus_tag="SF3531"
 /note="Residues 1 to 450 of 450 are 99 pct identical to
 residues 1 to 450 of a 450 aa protein from Escherichia
 coli O157:H7 ref: NP_312399.1"
 /codon_start=1
 /transl_table=11
 /product="glutathione oxidoreductase"
 /protein_id="AAN44987.1"
 /db_xref="GI:24053992"
 /translation="MTKHVDYIAIGSSGGIASINRAAYGKCALIBAKELGGTCVN
 VCGVKKVWMAAQI:REAHYMGPDYGTDTINKENWETLIASRTAYIDRLHTSVENY
 LCKNNVDVTKGPARVDKATLEVNGSTITADHILATGSRSHPDIPGVGIDSDGP

gene
 CDS
 67..1419
 /gene="gor"
 /locus_tag="SF3531"
 67..1419
 /gene="gor"
 /locus_tag="SF3531"
 /note="Residues 1 to 450 of 450 are 99 pct identical to
 residues 1 to 450 of a 450 aa protein from Escherichia
 coli O157:H7 ref: NP_312399.1"
 /codon_start=1
 /transl_table=11
 /product="glutathione oxidoreductase"
 /protein_id="AAN44987.1"
 /db_xref="GI:24053992"
 /translation="MTKHVDYIAIGSSGGIASINRAAYGKCALIBAKELGGTCVN
 VCGVKKVWMAAQI:REAHYMGPDYGTDTINKENWETLIASRTAYIDRLHTSVENY
 LCKNNVDVTKGPARVDKATLEVNGSTITADHILATGSRSHPDIPGVGIDSDGP

PALPALPERVAVGAGYIAVELAGVINGLGAKTHLFPVKHAPLRSFDPMISSETLVEM
 NAEGPOLHTNAIKPAVKNADGSLTLEDRSETVCLTIWAGREPANINLEAAG
 VTKNEGIYVDYQNTNIRGIVAGNDTGAVALTPVAAGARLSRLFNPKPDEHL
 DYSNIPVTSHPPIGTVGLTPOARQYGDQVVKYSFTAMYTAVTTRQPCRMK
 LVCVPKEIKVIGIHGIGFQDMLQGFVALKMGATKDDFNTVAIHPTAAEEFVTR
 complement (2332..2727)
 /gene="insB"
 /locus_tag="SF3532"
 complement (2332..2727)
 /gene="insB"
 /locus_tag="SF3532"
 /note="Residues 1 to 131 of 131 are 100 pct identical to
 residues 37 to 167 of a 167 aa protein from Shigella
 flexneri pir: T44475"
 /codon_start=1
 /transl_table=11
 /product="orf, conserved hypothetical protein"
 /protein_id="AAN44988.1"
 /db_xref="GI:24053994"
 /translation="MTVCAEMDEHMGVGAKSQRMLFYAYDRIRRTVAVFGERTL
 ATLERLLLSAFEVVVMWMDGNPLYESRLKGLHVISKRYTORIERHNLNLOHRL
 LVRKSUSFSKSVBLHDKVIGHYLNKHQ"
 complement (2754..3029)
 /gene="insA"
 /locus_tag="SF3533"
 complement (2754..3029)
 /gene="insA"
 /locus_tag="SF3533"
 /note="Residues 1 to 91 of 91 are 98 pct identical to
 residues 1 to 91 of a 91 aa protein from Shigella flexneri
 ref: NP_085394.1"
 /codon_start=1
 /transl_table=11
 /product="orf, conserved hypothetical protein"
 /protein_id="AAN44989.1"
 /db_xref="GI:24053995"
 /translation="MASISIRCPSCSATEGVNKGSTAGHQRYLCHCRKTWLOQT
 YTASQPTHOKIIDMAMNGVCASARIMGVGLTVLRHLKNSGRS"
 3071..3616
 /locus_tag="SF3534"
 3071..3616
 /locus_tag="SF3534"
 /note="no database matches"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAN44990.1"
 /db_xref="GI:24053996"
 /translation="MAALPCYPCNSKFLMKPVWASFYCIANWESVASALSPLS
 RSPRYPAIHCCVKAGCWTSKVSQSPITAYRIFQGRKRLMKPGDVNRKPROLS
 ATWPKTVPTVPTFAVKNLAKHINICTVLGRHYVTGRYLCPDRHRTGLAAERFR
 HRLECNARRGTGVGYGRGSSG"
 3588..4772
 /gene="arsB"
 /locus_tag="SF3535"
 3588..4772
 /gene="arsB"
 /locus_tag="SF3535"
 /note="Residues 1 to 394 of 394 are 100 pct identical to
 residues 36 to 429 of a 429 aa protein from Escherichia
 coli O157:H7 EDL933 ref: NP_290073.1"
 /codon_start=1
 /transl_table=11
 /product="arsenical pump membrane protein"
 /protein_id="AAN44991.1"
 /db_xref="GI:24053997"
 /translation="MVTGVVHPGDIPIVVMNIVMNATAFTAVIIISLLDESGFEWA
 ALHYSRWNGRGRILFTWIVLLGAVALPANDGAALITPIVIMLAILGFGKGTFL
 AFVMAQGITADTASLPLIVSNLVNISADFFGLGPRYASVMPVDIAIIVATVIMHL
 LYPRKQIPONTDMALKSPAEIKDPATRTGWWVLLLVGPFVLPGLGIPVSAIAA
 VGLILVPAKRGHAINTGKLGAPQIVTFSLGWLVVYVGLNAGLTYVLSGLNV
 LADNLGWAATLGTGTLTAFLSSINNMPTVLVGLSALDGSAGSVIKEMVYANVIC

```

Query Katch          100.0%;   Score 40;   DB 1;   Length 11524;
Best Local Similarity 100.0%;   Pred. NO. 0.0018;
Matches 40;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

QY      1  TAAGTCAAAACATATATGACTTAACGAATGTGTAAAGTC 40
DB      2253 TAAGTCAAAACATATATGACTTAACGAATGTGTAAAGTC 2214

RESULT 4
AC145934
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT

AC145934      179941 bp      DNA      linear      HTG 01-AUG-2003
Gallus gallus chromosome UNK clone CH261-4211, *** SEQUENCING IN
PROGRESSES ***, 44 unordered pieces.
AC145934.1  GI:33386884
HTG; HTGS_PHASE1.
Gallus gallus (chicken)
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
Wilson,R.K.
The sequence of Gallus gallus clone
1 (bases 1 to 179941)
2 (bases 1 to 179941)
Wilson,R.K.
Direct Submission
Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: JAA042101

* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1176: contig of 1176 bp in length
* 1177 1276: gap of unknown length
* 1277 2558: contig of 1282 bp in length
* 2559 2659: gap of unknown length
* 2659 3948: contig of 1190 bp in length
* 3949 5266: gap of unknown length
* 5266 5367: contig of 1318 bp in length
* 5367 6402: contig of 1035 bp in length
* 6402 6502: gap of unknown length
* 6502 7599: contig of 1098 bp in length
* 7599 7600: gap of unknown length
* 7600 8331: contig of 1132 bp in length
* 8331 8932: gap of unknown length
* 8932 10482: contig of 1550 bp in length
* 10482 10582: gap of unknown length
* 10582 12037: contig of 1456 bp in length
* 12037 12137: gap of unknown length
* 12137 13596: contig of 1459 bp in length
* 13596 13697: gap of unknown length
* 13697 14911: contig of 1215 bp in length
* 14911 15012: gap of unknown length
* 15012 16379: contig of 1268 bp in length
* 16379 16379: gap of unknown length

```

* 16380 17948: contig of 1569 bp in length
* 17949 18048: gap of unknown length
* 18049 19612: contig of 1564 bp in length
* 19613 19712: gap of unknown length
* 19713 21723: contig of 2011 bp in length
* 21724 21823: gap of unknown length
* 21824 25366: contig of 3443 bp in length
* 25367 25366: gap of unknown length
* 25367 27105: contig of 1739 bp in length
* 27106 27205: gap of unknown length
* 27206 28536: contig of 1331 bp in length
* 28537 28636: gap of unknown length
* 28637 30905: contig of 2269 bp in length
* 30906 31005: gap of unknown length
* 31006 33224: contig of 2219 bp in length
* 33225 33224: gap of unknown length
* 33225 37384: contig of 4060 bp in length
* 37385 37484: gap of unknown length
* 37485 40797: contig of 3313 bp in length
* 40798 40897: gap of unknown length
* 40898 43117: contig of 2220 bp in length
* 43118 43217: gap of unknown length
* 43218 46243: contig of 3028 bp in length
* 46244 46343: gap of unknown length
* 46344 49765: contig of 3422 bp in length
* 49766 49865: gap of unknown length
* 49866 56038: contig of 6173 bp in length
* 56039 56138: gap of unknown length
* 56139 60188: contig of 4050 bp in length
* 60189 60288: gap of unknown length
* 60289 67006: contig of 6718 bp in length
* 67007 67106: gap of unknown length
* 67107 68878: contig of 1772 bp in length
* 68879 68978: gap of unknown length
* 68979 72905: contig of 3927 bp in length
* 72906 73005: gap of unknown length
* 73006 77513: contig of 4508 bp in length
* 77514 77613: gap of unknown length
* 77614 81177: contig of 3564 bp in length
* 81178 81277: gap of unknown length
* 81278 86612: contig of 5335 bp in length
* 86613 86712: gap of unknown length
* 86713 91513: contig of 4801 bp in length
* 91514 91613: gap of unknown length
* 91614 97537: contig of 5924 bp in length
* 97538 103312: contig of 5675 bp in length
* 103313 103412: gap of unknown length
* 103413 111607: contig of 8195 bp in length
* 111608 123016: gap of unknown length
* 123017 123116: contig of 11309 bp in length
* 123117 131760: contig of 8644 bp in length
* 131761 131860: gap of unknown length
* 131861 133264: contig of 1404 bp in length
* 133265 133364: gap of unknown length
* 133365 140148: contig of 6784 bp in length
* 140149 140248: gap of unknown length
* 140249 152045: contig of 11797 bp in length
* 152046 152145: gap of unknown length
* 152146 165190: contig of 13045 bp in length
* 165191 165290: gap of unknown length
* 165291 179941: contig of 14651 bp in length.

FEATURES
source

1..179941 /organism="Callus gallus"
/mol_type="genomic DNA"
/db_xref="taxon:9031"
/chromosome="UNK"
/clone="CH261-4211"
1..1176 /note="assembly_name:Contig10"
1277..2558

misc_feature
misc_feature

misc_feature /note="assembly_name:Contig17"
2659..3848
misc_feature /note="assembly_name:Contig25"
3949..5266
misc_feature /note="assembly_name:Contig30"
5367..6401
misc_feature /note="assembly_name:Contig31"
6502..7599
misc_feature /note="assembly_name:Contig32"
7700..8831
misc_feature /note="assembly_name:Contig36"
8932..10481
misc_feature /note="assembly_name:Contig39"
10582..12037
misc_feature /note="assembly_name:Contig40"
12138..13596
misc_feature /note="assembly_name:Contig41"
13697..14911
misc_feature /note="assembly_name:Contig42"
15012..16279
misc_feature /note="assembly_name:Contig43"
16380..17948
misc_feature /note="assembly_name:Contig44"
18049..19612
misc_feature /note="assembly_name:Contig45"
19713..21723
misc_feature /note="assembly_name:Contig46"
21824..25266
misc_feature /note="assembly_name:Contig47"
25367..27105
misc_feature /note="assembly_name:Contig48"
27206..28536
misc_feature /note="assembly_name:Contig49"
28637..30905
misc_feature /note="assembly_name:Contig50"
31006..33224
misc_feature /note="assembly_name:Contig51"
33325..37384
misc_feature /note="assembly_name:Contig52"
37485..40797
misc_feature /note="assembly_name:Contig53"
40898..43117
misc_feature /note="assembly_name:Contig54"
43218..46243
misc_feature /note="assembly_name:Contig55"
46344..49765
misc_feature /note="assembly_name:Contig56"
49866..56038
misc_feature /note="assembly_name:Contig57"
56139..60188
misc_feature /note="assembly_name:Contig58"
60289..67006
misc_feature /note="assembly_name:Contig59"
67107..68878
misc_feature /note="assembly_name:Contig6"
68979..72905
misc_feature /note="assembly_name:Contig60"
73006..77513
misc_feature /note="assembly_name:Contig61"
77614..81177
misc_feature /note="assembly_name:Contig62"
81278..86612
misc_feature /note="assembly_name:Contig63"
86713..91513
misc_feature /note="assembly_name:Contig64"
91614..97537

Query Match 100.0%; Score 40; DB 2; length 179941;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAACTCAAAACATATATGACTTAACGATCTGTGAAGTC 40
|||||

```

Db      132260 TAAGTCARAAACATATATGACTTAACGAATGTGTAAAGTGC 132299

RESULT 5
ECOW76/c
LOCUS   ECOW76
DEFINITION E. coli chromosomal region from 76.0 to 81.5 minutes.
ACCESSION U00039
VERSION   U00039.1 GI:466582
KEYWORDS
SOURCE   Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1 (bases 1 to 225419)
Sofia,H.J., Burland,V., Daniels,D.L., Plunkett,G. III and
Blattner,F.R.
TITLE    Analysis of the Escherichia coli genome. V. DNA sequence of the
region from 76.0 to 81.5 minutes
Nucleic Acids Res. 22 (13), 2576-2586 (1994)
JOURNAL MEDLINE
PUBMED   94316500
2 (bases 1 to 225419)
Plunkett,G. III.
Direct Submission
Submitted (25-MAR-1994) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolgenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
COMMENT  This sequence was determined as part of the E. coli Genome Project
(Pfeiderick R. Blattner, director) at the University of
Wisconsin-Madison. Supported by award HG00301 from the NIH Human
Genome Project. The entire sequence was independently determined
from E. coli MG1655; overlaps and conflicts with other sequence
determinations are annotated. The start of this entry overlaps the
start of the entry ECOW82 (U01028) by 547 bp.
FEATURES             Location/Qualifiers
     source           1..225419
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K-12"
                     /db_xref="taxon:562"
                     /sub_strain="MG1655"
     notes            1..225419
                     /notes="This sequence comprises the following lambda
clones: EC19-98, EC27-273, EC17-131, EC30MM5, EC30K609-3,
EC27-334, EC30K606A-14, EC30K606A-12, EC30-154, EC27-385,
EC27-1129, EC30MM9, EC27-239, EC30MM3, EC27-SF1955,
EC18-200, EC27-1135, EC30R3900-8B, EC30R3900-5B, EC19MM6,
EC30K573-3; M13mp19 or Janus vectors were used for
subcloning"
     misc_feature     complement (<1..1221)
                     /notes="corresponds to M28722; ECOGGT (928..2148)"
     gene             complement (<1..1085)
                     /genes="ggt"
     CDS              complement (<1..1085)
                     /genes="ggt"
                     /note="95t" (<1..1085)
                     /note="CG Site No. 18280; cold-regulated"
                     /codon_start=1
                     /transl_table=11
                     /product="gamma-glutamyltranspeptidase"
                     /protein_id="AAB18422.1"
                     /db_xref="GI:466583"
                     /translation="MKPTPLRRVAIALISGSCPSAAAPAPVSVGVEDVPHFV
RAKQGMVASDATATQGVVDILKEGNAVDAAVAGYALAVTHPQAGNLGGGGFMLIR
SKNGNTTADIFREMAKATRDNFLDDQGNPDSSKSLTSLASGTGTGAGFSLADLK
YGTPLNKVQPKLARDGFIYNDALADDLTGYSEVLPHNENSKAI FWKGEPLKK
GDTLVQNLAKSLMEIAENGDPDFYKGTIAEQIAQEMKNGGLITKEDLAAKAVERT
PISGDRGVGVSMPPSPSGGIHIVQILNILENFDNMKKYGFSGSADAWQIAEAEKAY
ADRSVLGDDPDKVVPQALTNKAYKS"
                     <60..15171
     misc_feature     /notes="corresponds to lambda clone EC19-98"
     promoter         1067..1094

```

```

/promoter            complement (1085..1112)
                     /note="39%"
                     /note="43%"
/promoter            complement (1146..1173)
                     /note="45%"
/misc_feature        complement (1181..2402)
                     /note="corresponds to X56908; ECUGPQQ (1..1222)"
                     1205..1645
                     /note="alternate gene name yhbA"
                     /codon_start=1
                     /transl_table=11
                     /label="ORF_ol46a"
                     /protein_id="AAB18423.1"
                     /db_xref="GI:466584"
                     /translation="MKRLILITALLPFFVGFAGPINTLANPNQGOYIISQOQMOTQMQ
TQIQIQKGLNQQLTQLOQLQHLENNINNSQVLOSQPGERNPARQQLPNTNGG
MLNSNRNPDSGLNQQLPERRNGDMLNQPSPTQPDILKTI GP"
                     complement (1565..2383)
                     /note="corresponds to X14437; ECUGPQ (1..819)"
                     complement (1632..2375)
                     /gene="ugpQ"
                     complement (1632..2375)
                     /gene="ugpQ"
                     /EC_number="3.1.4.2"
                     /codon_start=1
                     /transl_table=11
                     /product="glycerophosphocholine phosphodiesterase"
                     /protein_id="AAB18424.1"
                     /db_xref="GI:466585"
                     /translation="MSNHPYPRIVARHGGKGLAPENTLASIDVGAKYGHMTEFDAKL
SKDGIFILLHDDNLESTNGWAGELNWDLLRVDAGSWSKMFKGEPLLSQVAE
RCRHRGWANTISIKPTTGTGKQVVALAARELWAGMTPLLSSFEIDALEAQAQAA
PRLPGLLLDEWDDMRBLTARLIGCVSLHNHKLANKARVMQLKDALRLIVTVNKP
QRAALLRWGVCICTDAIDVIGPNFTAQ"
terminator           1693..1727
                     /note="terminator-like sequence"
misc_structure       2192..2323
                     /note="predicted bend of 73.99 degrees"
misc_feature         complement (2225..6941)
                     /note="corresponds to X13141; ECUGP (1..4717)"
                     complement (2372..3481)
                     /gene="ugpC"
                     complement (2372..3481)
                     /gene="ugpC"
                     /function="sn-Glycerol-3-phosphate transport system"
                     /note="CG Site No. 17794"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB18425.1"
                     /db_xref="GI:912455"
                     /translation="MPSCAAWSIVRNKAGLKLQAVTKSWGKTQVTKPLTLDVADGE
FIVWVPSGCKSTLLRMVAGLRYVTEGDIWINDQVTEMEPKDRGLAMVFQNYALYP
HMSVEENMAGLIRGMCKQQLAERVKEAARILELDGLKKRPRELSGGQQRVAMGR
AIVRDPVPLDPEPLSNDAKLRVOMLELOQLHRLKLTSLVYVTHDOVEAMTLQVRV
WYMGGAQEQGTFTVEYVEKPSLFFVAFSGSPAMNLTGRVNNSTHPELGGIELP
LNGGYRQYAGKWTGIRPEHIALSSQAGGVFWMDLLEILGADNLAHGRWGEKLV
VRLAQHRTAGSTLMLHLAENQLHFDGRTGQRV"
                     complement (3444..4289)
                     /gene="ugpE"
                     complement (3444..4289)
                     /gene="ugpE"
                     /function="sn-Glycerol-3-phosphate transport system"
                     /note="CG Site No. 17791; membrane protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB18426.1"
                     /db_xref="GI:466587"
                     /translation="MIENRPMLTIFSHMLILIGAVILPPLVAVFAATLDKQAVVAA
PMTLIPGTHLENIHNVGVTNSAPFMRLNLSFVMAFSITLTKITVSLASAFAI
WVPRFPLNLPFWMIPIFILMPVEVRIPTVEVIANLQMLQVAGYGLPLMASATATP
LPRQFMILPDELVEAARIGASPMRFFCDIVFPLSKNLAAALFUTIFTYGNQWYLP
LLIITVDLGTTFVAGIKMIATGEGTETWNSVVMAMLLTLPPIPVVILVWQRAFVRGL
VDSEK"

```

```

gene      complement(4286..5173)
/genes="ugpA"
CDS       complement(4286..5173)
/genes="ugpA"
/function="sn-Glycerol-3-phosphate transport system"
/notes="CG Site No. 40"
/codon_start=1
/transl_table=11
/protein_id="AAB18427.1"
/db_xref="GI:466588"
/translation="MSSSRPVRSRWPLVLLVAPOLLITVIFINPAGNALWYSLQSV
DPFGSSQVGLDNFVTLPHDSYILDSFTTINKFTFTVVSGLLSLFFAALIVETIV
GSRFYTMLLPIYAPAAVAANIFLNPGRGLTHFLAEFGYDNNHQAQNSQAMFL
VVFASVTKQISYNFIFAAALQSIIRSLIEAAIDGAGFIRFFXIALPLTAPVSFEL
LVNLIYAFDFPFVVIDAATSGGPVQAATLIYKYREGFTGLDASSAAQSVLMFL
VVLTVQFRYVESKRYQ"
complement(5271..6587)
/genes="ugpB"
CDS       complement(5271..6587)
/function="sn-Glycerol-3-phosphate transport system"
/notes="CG Site No. 39; periplasmic binding protein"
/codon_start=1
/transl_table=11
/protein_id="AAB18428.1"
/db_xref="GI:466589"
/translation="MKPLRTASALALGALMNAQAVTIPFWHSMGELGKVEDSL
AQRFNAEDYKIVPTKYNGEONLSAGIAPRTGNAPALIOYIEVGTATMWSAKIK
PYDVFKEAGIQFQVPTGYSQSKTGHLSQFNSSTPYLYNKDAFKKAGL
DPEQPKWQDLADYAAKLSGCMKGVASGQWQIQLFNSAMNGLPASKNGPFG
DVALEFNKPEOVGMVADAKAPONAIIGASLWMOGKDETKFYNGCAMTASSGSLAN
IREYAKFNVGVMPPVDADAKAPONAIIGASLWMOGKDETKFYNGCAMTASSGSLAN
ENAEWHOKTGYLPIITKAAYDLTREGVEKNPGADATROMLNKGPPLPFTKGLNGLN
MQQIRVIVDEELSVWTKKTPQQALDTRVNGNQLRPFKSTKS"
misc_difference 5473..5474
/notes="GC in X13141; SS here"
misc_difference 5588
/notes="T in X13141; G here"
misc_difference 6021..6024
/notes="GGCG in X13141; SSGS here"
misc_feature   complement(6618..15331)
/notes="corresponds to J05516; ECOLIIVHMGF(1..8703)"
6640
/notes="mRNA start determined by S1 mapping; Molecular
Microbiol. 2:767 (1988)"
/evidence=experimental
promoter       complement(6651..6673)
/notes="includes pho box"
promoter       complement(6695..6723)
/notes="768"
repeat_region  6826..6859
/standard_name="REP; repetitive extragenic palindromic
element"
/notes="contains 1 REP sequence"
misc_difference 6836..6837
/notes="CG in J05516; GC in X13141 and here"
misc_difference 6845..6846
/notes="C in J05516; CC in X13141 and here"
complement(6986..7711)
gene          100.0%; Score 40; DB 1; Length 225419;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAGTCAAAAACATATATGACTTAACGAATGTGTAAGTGC 40
|||||
DB 62741 TAAGTCAAAAACATATATGACTTAACGAATGTGTAAGTGC 62702
|||||

RESULT 6
AC146183/c 242495 bp DNA linear HTG 04-NOV-2003
LOCUS Pan troglodytes chromosome y clone CH251-548L16, *** SEQUENCING IN
DEFINITION

```

```

PROGRESS ***, 6 unordered pieces.
AC146183
AC146183.2 GI:38154191
HTG; HTGS_PHASE1.
Pan troglodytes (chimpanzee)
Pan troglodytes
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Wilson,R.K.
1 (bases 1 to 242495)
The sequence of Pan troglodytes clone
Unpublished
2 (bases 1 to 242495)
Wilson,R.K.
Direct Submission
Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 242495)
Wilson,R.K.
Direct Submission
Submitted (04-NOV-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
On Nov 4, 2003 this sequence version replaced gi:33387136.

----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: bases at least Q40
Consensus quality: bases at least Q30
Consensus quality: bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1269: contig of 1269 bp in length
* 1270 1369: gap of unknown length
* 1370 2935: contig of 1566 bp in length
* 2936 3035: gap of unknown length
* 3036 4616: contig of 1581 bp in length
* 4617 4716: gap of unknown length
* 4717 6291: contig of 1575 bp in length
* 6292 8254: gap of unknown length
* 8255 8355: gap of unknown length
* 8355 242495: contig of 234141 bp in length.

FEATURES             Location/Qualifiers
     source            1..242495
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="Y"
                     /clone="CH251-548L16"
     misc_feature      1..1269
                     /note="assembly_name:Contig12"
     misc_feature      1370..2935
                     /note="assembly_name:Contig13"
     misc_feature      3036..4616
                     /note="assembly_name:Contig14"
     misc_feature      4717..6291
                     /note="assembly_name:Contig18"
     misc_feature      6392..8254
                     /note="assembly_name:Contig19"
     misc_feature      8355..242495
                     /note="assembly_name:Contig20"

ORIGIN

```

```

Query Match      100.0%; Score 40; DB 2; Length 242495;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAGTCAAAAACATATATGACTTTAAGCAATGCTTAAGTGC 40
Db 3353 TAAGTCAAAAACATATATGACTTTAAGCAATGCTTAAGTGC 3314

RESULT 7
AE016992 Shigella flexneri 2a str. 289816 bp DNA linear BCT 22-APR-2003
LOCUS Shigella flexneri 2a str. 2457T section 15 of 16 of the complete
DEFINITION genome
ACCESSION AE016992
VERSION AE016992.1 GI:30043426
SOURCE Shigella flexneri 2a str. 2457T
ORGANISM Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales;
          Enterobacteriaceae; Shigella.
REFERENCE 1 (bases 1 to 289816)
          Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
          Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
          Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
          Schwartz,D.C. and Blattner,F.R.
          Complete Genome Sequence and Comparative Genomics of Shigella
          flexneri Serotype 2a Strain 2457T
          Infect. Immun. 71 (5), 2775-2786 (2003)
          12704152
          2 (bases 1 to 289816)
          Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
          Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
          Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
          Schwartz,D.C. and Blattner,F.R.
          Direct Submission
          Submitted (13-JUN-2002) Genetics Laboratory, University of
          Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
          Location/Qualifiers
          1..289816
             /organism="Shigella flexneri 2a str. 2457T"
             /mol_type="genomic DNA"
             /strain="2457T"
             /serotype="2a"
             /db_xref="taxon:198215"
             complement(238..450)
             /gene="cspA"
             /locus_tag="S4179"
             complement(238..450)
             /gene="cspA"
             /locus_tag="S4179"
             /function="regulator; Adaptations, atypical conditions"
             /note="residues 1 to 70 of 70 are 100.00 pct identical to
             residues 1 to 70 of 70 from Escherichia coli K-12 : B3556"
             /codon_start=1
             /transl_table=11
             /product="cold shock protein 7.4, transcriptional
             activator of hns"
             /protein_id="AAP19147.1"
             /db_xref="GI:30043427"
             /translation="MSGKWTGLVXWFNADKGFITPDGSKDVVFHFSATQNDGYKS
             LDEGQVSTIESGAKGPAAGNVTSL"
             complement(896..1021)
             /gene="yiaG"
             /locus_tag="S4180"
             complement(896..1021)
             /gene="yiaG"
             /locus_tag="S4180"
             /note="residues 1 to 41 of 41 are 100.00 pct identical to
             residues 1 to 41 of 96 from Escherichia coli K-12 : B3555"
             /codon_start=1
             /transl_table=11

gene
/product="hypothetical protein"
/protein_id="AAP19148.1"
/db_xref="GI:30043428"
/translation="MEYKDPHELLSSLEQIVFNDETQKILTHRETTTCTETQL"
1335..2165
/gene="yiaP"
/locus_tag="S4181"
1335..2165
/gene="yiaP"
/locus_tag="S4181"
/note="residues 1 to 276 of 276 are 94.56 pct identical to
residues 1 to 276 of 276 from Escherichia coli K-12 :
B3554"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP19149.1"
/db_xref="GI:30043429"
/translation="MILPGRRLKILQACPLSLRQTRVCRCLFLGERSKKMATG
KSCRWNPALALLMVSLSCFKEGDKQKAFIDFLONTWRSGERLPTLTADQKKQ
FGPVSVDYALIGYSQQVQNMDSGLRPVDSVNAIRVPDIYVITQSGLRENGSLGV
LAQQLQNAKLQADAHSALXQSDDLKPFVDQFTKVVTTPADALQPLIPAAQTFQQL
VMVGDTAAQQTQVSFVANGIQFPTSQQASEYNKLIAPLPAQHAQFNAQWATVATATQ
"
complement(2215..3189)
/gene="yiaP"
/locus_tag="S4182"
complement(2215..3189)
/gene="yiaP"
/locus_tag="S4182"
/function="putative enzyme; Not classified"
/note="residues 1 to 324 of 324 are 94.13 pct identical to
residues 5 to 328 of 328 from Escherichia coli K-12 :
B3553"
/codon_start=1
/transl_table=11
/product="putative dehydrogenase"
/protein_id="AAP19150.1"
/db_xref="GI:30043430"
/translation="MKPSVILYKALPDDLQRLQAHFTVHOVANLSPTVQVNAIPA
EAGLGSNENVDAALEKMKLRATSTISVGNDFVDALTSKILLMHPVTLTET
VADTLMAIVLSTARVVEAERVKAGETASIGDPDWYGTVDVHHKTLGVGCRIGMAL
AQRVHFGNMPILTNHRRHKEERFNARFCDDTLIQESDFVCLILPLDTEHLLF
GAQFQAKMSAIPINAGRPVVDENALIALQKEITHAAGLDVFEQEPUSVDSPLLS
MANVAVPHIGSATHTETRYGMAACAVDNLDLALQKVEKNCVNPVAD"
complement(3293..3919)
/gene="yiaD"
/locus_tag="S4183"
complement(3293..3919)
/gene="yiaD"
/locus_tag="S4183"
/function="putative membrane; Not classified"
/note="residues 1 to 208 of 208 are 88.94 pct identical to
residues 12 to 219 of 219 from Escherichia coli K-12 :
B3552"
/codon_start=1
/transl_table=11
/product="putative outer membrane protein"
/protein_id="AAP19151.1"
/db_xref="GI:30043431"
/translation="MSGALVSGCTTNPYTGGERAGKSAICAGISLVGAGIGALSSS
KDRGKGAALIGAAAGALGGVGYMDQVKLRDKMRGTGVSVTRSDNILLNPNPN
VTDTSSAPLKPAGANTLTGVAVMLKEYPKTAIVNYIGYTDSTGCHDLNMRLSQQRADS
VASALITQGVDSKRITQGLGPANPIASNSTAGKAQNRNRVEITLSEI"
4160..6439
/gene="bisc"
/locus_tag="S4184"
4160..6439
/gene="bisc"
/locus_tag="S4184"
/function="enzyme; Biosynthesis of cofactors, carriers:
Biotin"
/note="residues 21 to 759 of 759 are 98.78 pct identical

```

to residues 1 to 739 of 739 from Escherichia coli K-12 :
B3551"
/codon_start=1
/transl_table=11
/product="biotin sulfoxide reductase"
/protein_id="AAP19152.1"
/db_xref="GI:30043432"
/translation="MLVETDGETVSSRGALATGKNSLQSAVRDQVHSNTRVRPPMV
RKFPLASPNPGIQRGDEFVRSWDEALDHLHQHKKRIEAYGPASIFAGSYGRSN
GVHLKASLLQRYMALAGYTGHLGDYTGAAQAIIMPVVGSEYVQOQTSPWLVEH
SDVVLWGANPLNTLKIWNASDEQGLSYFSAIRDSGKKLICIDPMRSETVDFPGDKM
SWAPFMGTDLVLMGIARTLVENGHDEAFIARCTTGVAVPASVLLGESDGIARNAE
WAAEICGVGAHRIELAI PHONTMLMAGMQRQFGEOKHMIIVTLAAMCGIGT
PGGGFGLFVANGNPTRRSAVLSSMOGSLPGGCDVVKIPVARIIVALENRCGAYQ
ENKMRHPDIFPIWAGGNPTHQDTHNLRKQKDELVIIVISCFWTAARADIV
LPATTSFERNLMTGTGDSNQLVKNQVPPFYEARNDFFVIELSERWEKGFARF
TEGSQLQWLETFTVNAQRGASQVQVFPFABFQANQLIEMFENPDSRFIRFAD
CRDPLAHLKTASGKIEIFSGRIADYPCDGFHMLPEDEMQRNAPQLOVLSAH
FAHLHSQNTSSSRELYAVANREPTTHPDQAQGITGDMVVRWNSRQIILAGAV
ISEGKPGVICIHEGAWPDLOLTADGICKNGAVNVLTKDPSRLGNSCAGNTTLAWL
EYKNGPFLTLTAPEPPASS"
complement(6408..6848)
/gene="yiaC"
/locus_tag="S4185"
complement(6408..6848)
/gene="yiaC"
/locus_tag="S4185"
/note="residues 1 to 146 of 146 are 97.94 pct identical to
residues 1 to 146 of 146 from Escherichia coli K-12 :
B3550"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP19153.1"
/db_xref="GI:30043433"
/translation="MIRAEORSELPAILLEWLESTTGHPPFKANYWEDCIPVVRDAY
LANAQWYEDGKLGKLVGSIMEGRFPLAAMVAPKAVRGRIGKALMQVQRYPHMLM
EYVQKQPAIDFYAQGFHIVDCAWQDETQTPWMSFVWQTIL"
complement(6845..7408)
/gene="tag"
/locus_tag="S4186"
complement(6845..7408)
/gene="tag"
/locus_tag="S4186"
/function="enzyme; DNA - replication, repair,
restriction/modification"
/note="residues 1 to 187 of 187 are 97.32 pct identical to
residues 1 to 187 of 187 from Escherichia coli K-12 :
B3549"
/codon_start=1
/transl_table=11
/product="3-methyl-adenine DNA glycosylase I"
/protein_id="AAP19154.1"
/db_xref="GI:30043434"
/translation="MERCGWVSGQPLXIYHDNHWGVPETDSKKLFEMICFEGQOAGL
SWITLKKRNYRAYFHQDPKVAAMQEDVERLVODAGIIRHRGKIQAIIGNRAY
LQWQNGEPFPDFVMSFNHQPQVTOATTLSEITPTSASDALSALKRGRFVGTIT
ICYSFMOACGLVNDHVHVCCLYGNKP"
7566..8264
/gene="yhjY"
/locus_tag="S4187"
7566..8264
/gene="yhjY"
/locus_tag="S4187"
/function="purative enzyme; Not classified"
/note="residues 1 to 232 of 232 are 99.13 pct identical to
residues 1 to 232 of 232 from Escherichia coli K-12 :
B3548"
/codon_start=1
/transl_table=11
/product="3-methyl-adenine DNA glycosylase I"
/protein_id="AAP19155.1"
/db_xref="GI:30043435"
/translation="MERCGWVSGQPLXIYHDNHWGVPETDSKKLFEMICFEGQOAGL
SWITLKKRNYRAYFHQDPKVAAMQEDVERLVODAGIIRHRGKIQAIIGNRAY
LQWQNGEPFPDFVMSFNHQPQVTOATTLSEITPTSASDALSALKRGRFVGTIT
ICYSFMOACGLVNDHVHVCCLYGNKP"
7566..8264
/gene="yhjY"
/locus_tag="S4187"
7566..8264
/gene="yhjY"
/locus_tag="S4187"
/function="purative enzyme; Not classified"
/note="residues 1 to 232 of 232 are 99.13 pct identical to
residues 1 to 232 of 232 from Escherichia coli K-12 :
B3548"

Query Match

Best Local Similarity 100.0%; Score 40; DB 1; Length 289816;

Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TAAGTCAAAAACATATGACTTAACGAATGTGTAAGTGC 40

|||||

Db 66023 TAAGTCAAAAACATATGACTTAACGAATGTGTAAGTGC 66062

RESULT 8

AE005575/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

AE005575 11071 bp DNA linear BCT 21-MAR-2001
Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 194
of 290.

AE005575 AE005174

AE005575.1 GI:12518196

Escherichia coli O157:H7 EDL933

Escherichia coli O157:H7 EDL933

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Escherichia.

1 (bases 1 to 11071)

Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,

Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,

Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,

Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,B., Potamousis,K.,

Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,

Weich,R.A. and Blattner,F.R.

Genome sequence of enterohaemorrhagic Escherichia coli O157:H7

Nature 409 (6819), 529-533 (2001)

21074935

11206551

2 (bases 1 to 11071)

Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,

Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,

Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,

Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,B., Potamousis,K.,

Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,

Weich,R.A. and Blattner,F.R.

Direct Submission

Submitted (22-OCT-2000) Laboratory of Genetics, University of

Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

Location/Qualifiers

1. 11071

/organism="Escherichia coli O157:H7 EDL933"

/mol_type="genomic DNA"

/strain="EDL933"

/serotype="O157:H7"

/db_xref="taxon:155864"

/note="enterohaemorrhagic"

67. 1419

/gene="gor"

/note="synonym: Z4900"

67. 1419

/gene="gor"

/function="enzyme; Biosynthesis of cofactors, carriers;

Thioredoxin, glutaredoxin, glutathione"

/note="Residues 1 to 450 of 450 are 99.33 pct identical to

residues 1 to 450 of 450 from Escherichia coli K-12 Strain

MG1655: B3500"

/codon_start=1

/transl_table=11

/product="glutathione oxidoreductase"

/protein_id="AGS5832.1"

/db_xref="GI:12518197"

/translation="MTKHYYIAIGGGGGIASINRAAMYQKCALIEARLGGTCVN

VGCVPKVMHAAQIREATHMTGPDYFTTINKFNWETLIASRTAYIDRIETSYENV

LGNRVDTKGPAPFADKLEVNGETITADHILIAIGKPSHPDIPGVYSDSGF

FALPALPVRVAVGAGYIAVLAVGKATLPRKAPLSPDPMISETLVSM

NAEGPOLHTNALPKAVKNADGSLTLEEDGSETVDCIWAIGREPANDNLNLAAG

VTNKGCHLVNDKYNVTVEGIYAGVDNTGAVELTFVAAGRLSERLNNKPDHIL

DYSNPTVYFHPPTVGLTFPQAREQYGDQVRYKSSFTAMYTAVTTHRQPCRMK

XVCVGPKEKIVGIGIGFQMDLQGFVAVLKKGATKDFDNTVAIHPTAAEFVTMR

2300..2653

/gene="arsR"

/note="synonym: Z4903"

2300..2653

/gene="arsR"

gene

CDS

```

/function="regulator; Protection responses: Drug/analog
sensitivity"
/notes="Residues 1 to 117 of 117 are 94.87 pct identical to
residues 1 to 117 of 117 from Escherichia coli K-12 Strain
MG1655: B3501"
/codon_start=1
/transl_table=11
/product="transcriptional repressor of chromosomal ars
operon"
/protein_id="AAG58633.1"
/db_xref="GI:12518198"
/translation="MSFLPIQLFKILADETRIGLIVLISELGEICVDCDLCTALDQSQ
PKISRLALLRESGLLDRKQKGVHYRLSPHPSWAAXIIIEQAWRCQEKKVQIVRN
LARQNSVDSKNTCS"
2707..3996
/gene="arsB"
/notes="synonym: Z4904"
2707..3996
/gene="arsB"
/function="transport; Protection responses: Drug/analog
sensitivity"
/notes="Residues 1 to 429 of 429 are 100.00 pct identical
to residues 8 to 436 of 436 from Escherichia coli K-12
Strain MG1655: B3502"
/codon_start=1
/transl_table=11
/product="arsenical pump membrane protein"
/protein_id="AAG58634.1"
/db_xref="GI:12518199"
/translation="MLLAGAIFVITIVIVIQPKLGIGKSNITGCAVLAVTVGVHPG
DIPVWNLVWNAAPAIPIVILISELSDSGFFFEAAALVSRMGNGRLLFTWVLLG
AAVALADFGDALITPTIVIAMLALGFSKGTTLAPYMAAGFIADTASLPLIVSNLV
NIVSADPFLGFRBYAVWVPVDIAIIVATLVMLHLYFRKDIPOYDMALLKSPAEAI
KDPATFKTGVVLLVLLGFFVLEPLGIPVSAIAVAGALILFVVAKRGHGINTGVLR
GAPWQIVTFSGLWLVVYGRNAGLTGYLSGVNLADMLMAATLGTGTAFPLSSI
KNNMPTVLVGLSIDGSTASGVIEAMVYANVIGDLPKPTPIQSLATLLWLVLSQ
KNNITSGVYFRTGINTLPVLFTVTLAALALSLTL"
4009..4434
/gene="arsC"
/notes="synonym: Z4905"
4009..4434
/gene="arsC"
/function="enzyme; Protection responses: Drug/analog
sensitivity"
/notes="Residues 1 to 141 of 141 are 98.58 pct identical to
residues 1 to 141 of 141 from Escherichia coli K-12 Strain
MG1655: B3503"
/codon_start=1
/transl_table=11
/product="arsenate reductase"
/protein_id="AAG58635.1"
/db_xref="GI:12518200"
/translation="MSNITIYHNACGTSRNTLEMRNSGTEPTIIHYLTPPTDEL
VKLIADMGISVRLARKVPEYEEGLAEKFDLDRLLDFMLQHPILINRPIVVTLLG
TRLRPSVRLPILPDAQGAFTKEDGEKVVDEAGRLK"
4546..4653
/gene="Z4906"
4546..4653
/gene="Z4906"
/function="orf; Unknown function"
/notes="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG58636.1"
/db_xref="GI:12518201"
/translation="MLFVLGTGWAHQAILLFTYSLRLEIPIIASSMMYQ"
5062..6177
/gene="Z4907"
5062..6177
/gene="Z4907"
/function="orf; Unknown function"
/notes="Residues 1 to 249 of 371 are 89.95 pct identical to
residues 1 to 249 of 260 from Escherichia coli K-12 Strain

```

```

MG1655: B3504"
/codon_start=1
/transl_table=11
/protein_id="AAG58637.1"
/db_xref="GI:12518202"
/translation="MSIDFTPGMINTYRGDIYNTTDTDNVKTPTDTPMPCDNREBQQ
PINTFSGEGYNPEQDLAROHLPQIVACRTNTKYTDADYSKVAQLVSLITNTESIS
STPTLROAQEILQINNIIRVEKKSACRIIVVANPKPKDALKITKISVEEGIPVRFVS
QTFMSDTNFIARQADLPNTIKDQSLQYQMTKLYIESENKMKVFAGNFIDFNM
TQNLGSPVLTLSRPFEDNLNINFTDANGDITFSSKHEHTPKLYKDEQYLDKQIEG
LFTSLTINDNFLRAKAIASITIIKPLEARITNLSYDXILKYQOEFQEKCYQKVSFT
LFTSYNIQITWAMSEYQFEVFTYETLNPKKCHARLI"
6499..7098
/gene="slp"
/notes="synonym: Z4908"
6499..7098
/gene="slp"
/function="membrane; Cell envelop: Outer membrane
constituents"
/notes="Residues 1 to 199 of 199 are 99.49 pct identical to
residues 1 to 199 of 199 from Escherichia coli K-12 Strain
MG1655: B3506"
/codon_start=1
/transl_table=11
/product="outer membrane protein induced after carbon
starvation"
/protein_id="AAG58638.1"
/db_xref="GI:12518203"
/translation="MWLFYVDKDSMMWMTKALILSLFLLAACSSIPQNIKENNQPD
IQKSFVAHQPGGLYVQCQAPGKVINVGKTDITLLEIAVLPLDSVAKPDIEANYQ
GRLLRASGDFDPVNYNHFVITILGTIQGEPGFINKVPVLEVMNGIOIQVHLLREV
VNTTINWDYGYGAFWPEPGWGAPEYTNVAVSQVTFELVK"
7254..7784
/gene="yhiP"
/notes="synonym: Z4909"
7254..7784
/gene="yhiP"
/function="orf; Unknown function"
/notes="Residues 1 to 176 of 176 are 99.43 pct identical to
residues 1 to 176 of 176 from Escherichia coli K-12 Strain
MG1655: B3507"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG58639.1"
/db_xref="GI:12518204"
/translation="MFLIITRDYMTFMKNIKSGNVHVIQNEEEDVWMLHQNAFVI
IDTLNNVPHSNFLTOIERLKPVHVIIIESPNIKRCIGKVPVTFVPTITIIDFVALI
NGSYCVPEANVSLSRKQHVLSICANQWTTEDILEKLSIKTFYCHKENIMMLNL
KRINELVRHOHIDVLY"
7786..>11071
/notes="O-Island #140; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
/complement(7845..8873)
/gene="chuS"
/notes="synonym: Z4910"
/complement(7845..8873)
/gene="chuS"
/function="putative transport; Transport of small
Query Match 88.0%; Score 35.2; DB 1; Length 11071;
Best Local Similarity 92.5%; Pred. No. 0.072;
Matches 37; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 TAACTCAAAACATATATGCTTAAACGAATGCTGAATGTC 40
|||||
Db 2254 TAACTCAAAACATATATGCTTAAACGAATGCTGAATGTC 2215
|||||
RESULT 9
AP002565/c 267888 bp DNA linear BCT 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 16/20.
DEFINITION

```


ACCESSION	AP002565	BA000007		/codon_start=1
VERSION	AP002565.1	GI:13363693		/evidence=not experimental
KEYWORDS				/transl_table=11
SOURCE	Escherichia coli O157:H7			/product="hypothetical protein"
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.			/protein_id="BAB37643.1"
REFERENCE	1 (sites)			/db_xref="GI:13363693"
AUTHORS	Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T., Yanamoto,K., Ohnishi,M., Hayaashi,T., Yasunaga,T., Honda,T., Sasaki,K., and Shinagawa,H.	gene		/translation="MNKFKVALVGAVLATLTACTGHIENRDKNCSDYDLLHPAISIS KIIGGCGPTAQ"
TITLE	Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak	CDS		563. .1951
JOURNAL	Genes Genet. Syst. 74 (5), 227-239 (1999)			/gene="ECs4221"
MEDLINE	20198780			/gene="ECs4221"
PUBMED	10734605			/note="similar to YHFM_ECOLI gi 2367216 percent identity 99 in 462 aa (Conserved in E.coli K-12)"
AUTHORS	Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M., Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and Hayashi,T.			/codon_start=1
TITLE	Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655			/evidence=not experimental
JOURNAL	Syst. Appl. Microbiol. 23 (3), 315-324 (2000)			/transl_table=11
MEDLINE	20557356			/product="putative amino acid/amine transport protein"
PUBMED	11108008			/protein_id="BAB37644.1"
AUTHORS	Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S., Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T., Yanamoto,K., Hayaashi,T., Yasunaga,T., Honda,T., Sasaki,K. and Shinagawa,H.			/db_xref="GI:13363695"
TITLE	Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak	gene		/translation="MTANSPQVRGQEKIAMSQBLORLKLGFVAVLAIAGTVTCSG IFVSGEVAKAAGTPTLVLAIFVIGLIVIPQMCVYAEULSTAPENGADTVYLNAGS RPLAFSGWASFWANDAPSLIMALAIVSNLGLTPIIDPLGKPIAGLIIAFMLLHL RSVEGGAFTQTLITIAKIIPFTIVIGLIFWPKAENFAAPATTAGTSGFMALLAGI SATSWSYTGNASICYMTGEIKNPKTMPRALIGSCLLVLYTLTLLAVISGLMPDFKL ANSETPTSDALTWIPALGSTAGIFVAITAMIVILGSLSCVMYQPRLEYAMAKDNLFK KCRGHVHPKYNTPDVSIILLOGALGIRFIVSDLTSLGYETLWMCFKNTLTFGSIIWC RKRDYKPLWRTPAFLMTPLAASSLILVASTFVWAPIPGLICAVIVATGUPAYAF WAKRSRLNALS"
JOURNAL	Gene 258 (1-2), 127-139 (2000)	CDS		1951. .2994
MEDLINE	20564182			/gene="ECs4222"
PUBMED	11111050			1951. .2994
AUTHORS	Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasaki,K., Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and Shinagawa,H.			/gene="ECs4222"
TITLE	Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12			/note="similar to YHFN_ECOLI gi 1789772 percent identity 98 in 347 aa (Conserved in E.coli K-12)"
JOURNAL	DNA Res. 8 (1), 11-22 (2001)			/codon_start=1
MEDLINE	21156231			/evidence=not experimental
PUBMED	11258796			/transl_table=11
AUTHORS	Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasaki,K., Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and Shinagawa,H.			/product="putative transport protein"
TITLE	Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12	gene		/protein_id="BAB37645.1"
JOURNAL	DNA Res. 8 (1), 11-22 (2001)	CDS		/db_xref="GI:13363696"
MEDLINE	21156231			/translation="MLSGENKMLDIDKSTVDVLVTENMVOEKEVLSHDVPLVHTIVE EMWRDLIRIFYACGSPNNAOAKHLADRFSDLQVYALSCHWECDCDTPYRLDARCA VIGVSDYKTEEVIKALELGRACGAATAFKAQSDSPITSAAEFSIDYQACDIWEIHL LLYGVVLEMITRLAPHAETGKIKNDLQFNALGYLVRTWEEKRGQLGASQWPMIL YTVAAQPLRPLGYKEGIVTLMFTWTHGCVESGEPRHGLEIPEVGPVFLFLGNDE SRUTRAINFVKQRTDNVIDVAAISQGLHFWLAPFLFVFPVMEMLCYLISYIKDHN PDERRYVYGLVEY"
PUBMED	11258796			3045. .3872
AUTHORS	Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and Hayashi,T.			/gene="ECs4223"
TITLE	Direct Submission			/gene="ECs4223"
JOURNAL	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:kengen-info.osaka-u.ac.jp, URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047)			/note="N-terminal part (1-115 in 294 aa) is similar to YHFO_ECOLI gi 1789773 percent identity 97 in 115 aa, C-terminal part (162-294 in 294 aa) is similar to YHPP_ECOLI gi 2367217 percent identity 99 in 133 aa"
COMMENT	genome project.			/codon_start=1
FEATURES	Location/Qualifiers			/evidence=not experimental
source	1. .267898			/transl_table=11
	/organism="Escherichia coli O157:H7"			/product="hypothetical protein"
	/mol_type="genomic DNA"			/protein_id="BAB37646.1"
	/strain="O157:H7"			/db_xref="GI:13363697"
	/sub_strain="RIMD 0509952"			/translation="MKTGMTTCGHORLPIEHAPDASELGYDGIIEWGRPHAPAPDL KAGGIKQIKALACTYQNPILGYTPETNGYPYNNMGLDHEMRBSLMDIKLAMDWAKEM NAGTTLISAAHAGHLPENVWIRLAENLSELEYAENIGMDLILELPTYESVAVCN ANDVLHALAVPSPRLFSWVDICAPYQAPFVMSYFDKGLGKHLHLIIVSDGASDTH YIPGEGMPLREBLMRDIIIDRGYGYCTVELVTWYMNPEFLYARQALRRFRALLPEDE"
	/db_xref="taxon:83334"	gene		3872. .4204
	151. .318			/gene="ECs4224"
	/gene="ECs4220"	CDS		3872. .4204
	131. .318			/gene="ECs4224"
	/gene="ECs4220"			/note="similar to N-terminal part (1-110 in 261 aa) of YHFO_ECOLI gi 1789775 percent identity 98 in 110 aa"

Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamaze, R., Landers, T., Lehoczy, J., Levine, R., Matthews, C., MacLean, C., MacDonald, P., Major, J., Marquis, N., Mattheus, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 166112)

Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, O., Lui, A., Mabbitt, R., MacLean, C., MacDonald, P., Major, J., Manning, J., Mattheus, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (04-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 4, 2003 this sequence version replaced gi:22381167.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE JOURNAL COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L18273
Center clone name: 200.D.15
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165601 bases at least Q40
Consensus quality: 165840 bases at least Q30
Consensus quality: 165983 bases at least Q20
Insert size: 158000; agarose-fp
Insert size: 165912; sum-of-ctnigs
Quality coverage: 12.6 in Q20 bases; agarose-fp
Quality coverage: 12.0 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 37442: contig of 37442 bp in length
* 37443 37542: gap of 100 bp
* 37543 101600: contig of 64058 bp in length
* 101601 101700: gap of 100 bp
* 101701 166112: contig of 64412 bp in length.

FEATURES source

1..166112
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-200D15"
/clone_lib="RPCI-24 Male Mouse BAC"
1..37442
/note="assembly_fragment
clone end:SP6
vector side:left"
37543..101600
/note="assembly_fragment"
101701..166112
/note="assembly_fragment
clone end:T7
vector side:right"

misc_feature

1..37442

/note="assembly_fragment
clone end:SP6
vector side:left"

misc_feature

37543..101600

/note="assembly_fragment"

misc_feature

101701..166112

/note="assembly_fragment
clone end:T7
vector side:right"

ORIGIN

Query Match 61.5%; Score 24.6; DB 2; Length 166112;
Best Local Similarity 76.9%; Pred. No. 1.5e+02;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 TAAGTCAAAACATATATGACCTTACGAAATGTGAAGTG 39
|||||
Db 76957 TAAGTCAAAACATATTTTGAATTCCTAATATGTAAGTG 76995
|||||

RESULT 12 AL772194/c

LOCUS
DEFINITION
AL772194 Mouse DNA sequence from clone RP23-331M18 on chromosome 4, complete sequence.
AL772194.6 GI:25136989

VERSION
AL772194.6

KEYWORDS
HTG.

SOURCE
Mus musculus (house mouse)

ORGANISM
Mus musculus

REFERENCE
1 (bases 1 to 176496)

AUTHORS
Harrison, B.

TITLE
Direct Submission

JOURNAL
Submitted (14-SEP-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 19, 2002 this sequence version replaced gi:22798144.

COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep RP23-331M18 is from the RPI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTER: pBACE3.6.
Location/Qualifiers
1..176496
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-331M18"
/clone_lib="RPI-23"

ORIGIN
Query Match 61.5%; Score 24.6; DB 10; Length 176496;
Best Local Similarity 76.9%; Pred. No. 1.5e+02;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 TAAGTCAAAACATATGACCTTACGACGTAAGT 39
|||||
Db 22991 TAAGTCAAAACATATGACCTTACGACGTAAGT 22953
|||||

RESULT 13
LOCUS AF242431S2 90650 bp DNA linear ROD 18-SEP-2000
DEFINITION Mus musculus Naip3 gene, exon 1; neuronal apoptosis inhibitory protein 1 (Naip1) and general transcription factor I1H polypeptide 2 (Gtf2h2) genes, complete cds.
ACCESSION AF242432
VERSION AF242432.1 GI:9082149
KEYWORDS 2 of 2
SEGMENT
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (Bases 1 to 90650)
Endrizzi,M.G., Hadinoto,V., Growney,J.D., Miller,W. and Dietrich,W.F.
Genomic sequence analysis of the mouse Naip gene array
Genome Res. 10 (8), 1095-1102 (2000)
20414747
PUBMED 10958627
REFERENCE 2 (Bases 1 to 90650)
Dietrich,W.F.
Direct Submission
TITLE Submitted (08-MAR-2000) Genetics, HHMI/Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA
JOURNAL
Location/Qualifiers
1..90650
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="13"
/clone="BAC 26f17"
/note="centromere distal sequence fragment"
complement(2254..2328)
/gene="Naip3"
/note="region of sequence similarity; not known to be transcribed"
/number=1
complement(6546..51581)

MRNA
/gene="Naip1"
/note="synonym: Naip, rs 7"
complement(join(6546..7941,9650..9814,12413..12499,12789..12941,21794..21961,24127..26238,26702..26759,27914..27995,31151..31250,31944..32063,35161..35212,36038..36119,42341..42440,42892..43482,45016..45057,51508..51581))
/product="neural apoptosis inhibitory protein 1"
complement(join(7577..7941,9650..9814,12413..12499,12789..12941,21794..21961,24127..26238,26702..26759,27914..27995,31151..31250,31944..32063,35161..35212,36038..36119,42341..42440,42892..43459))
/gene="Naip1"
/codon_start=1
/product="neural apoptosis inhibitory protein 1"
/protein_id="AA082752.1"
/db_xref="GI:9082151"
/translation="MAEHGSSSDRISEIDYFELPESALLGVDAVOLAKSQBEEHK
ERMKKGNQMSRAKRLTKFTYDFRSWTPQEMAAAGFYHTGVLGVCPCCSL
ILFNSLRKLPTRHKKLRPECFLOKQKNGIKYDIKVKSEPKMLGKARHYEE
ARLESDFWFFAHGTSRVLGSAAGFTKRTVQCFSCGSLGNEEGDDPWKEHA
KWFPCFELQSKSSSEIAQYIOYEGFVHTGEFVNSWRLELPSVAYCNDVFA
NEELMDTFKDPHESFVADALVRAGLFYTGKKIVQFCGCKMEKTEGDDPQIE
HNKFPNCIFLOTFKSAEVIIPALLSCALPEMETTESNDDPAHVTVGLGRS
EAQMFQARSLSQLRDNVTKATFRMNLPEVCSLGDHLICDVSIIISKHISQYQ
GALTIPVFNLSVMCEGETSGKTFKRIAFIWAAGCCPLVRFOLVFFVLSLS
ITPDGLANTICAILGAGGCISEVLSIIQQLGCVLFLLDPYGLASLPAALHTL
ITNYSRLTLLAVHTRVGRISVLDTSLEIKKPSLNTVYLKFPESHNTKELLE
FWYFGNEDLQGHKIPFVAAVCTDFNPSDQFQDMALFKSMYLSLXKKGAA
KPLQATVSCGQALTLGFPSSCFEFDSDDAEAGVDEDELTCLMSTYNNFLRPSYH
RFLGPLFQEFLEAAMRUTLELLSDRQEDQDLGYLRQINSPLKALTYNNFLRPSYH
PSSKAGPTVVSHLLHLVDETELLNTYKEDYVNHPPGTRIMKGLKMLLSPEYS
SPVSEHLRLIALNFAYSNVAVCSPILOFLGRITLALKVNLQVFRDHPESLLAVK
SVSINGNPKPVVDVSVMEKSFETLOPTIDQYASAFQMKHEKLSNEETIK
SIKNIPLQPKTSKCYKLKPKCKIPRLVGVGTMMGADQALLVLMHVPASOSI
RFLSDSSSFLESIRPALELSKASVTKCSMRLELSRAQELLITLPAQSLVSETN
QLPEQFHNHFKFGLKRCVDRGNVSLVLPREFPNLLHKEKLSIRTFPEMLDPSK
LVKLIQNSNLHFLKDLPLNSCDLMAVLACKKLEIEFSGRCFEAMPTFVNLIPN
PCLRLVLFYILDNDSVIETARVATSGGFKLEKLDLSNMHKTIEGYNRFFQALDNL
PNLQNLICRHIPCEIQVQATTVKALGQCVSLPSLTLHMLSWLDEEDMKVINDVK
ERHPSKRLIIFKMWIVFSPVLE"
12475..12811
/note="D13Die31"
31849..32057
/note="D13Die3"
complement(<67642..>88844)
/gene="Gtf2h2"
complement(join(<67642..67761,69615..69654,75206..75308,75841..75947,77928..77991,79262..79369,79726..79813,79934..80024,81662..81767,84790..84844,85193..85243,87569..87655,87856..87892,88202..88279,88789..88844))
/gene="Gtf2h2"
/product="general transcription factor I1H polypeptide 2"
complement(join(67642..67761,69615..69654,75206..75308,75841..75947,77928..77991,79262..79369,79726..79813,79934..80024,81662..81767,84790..84844,85193..85243,87569..87655,87856..87892,88202..88279,88789..88844))
/gene="Gtf2h2"
/note="Btf2; P44; coding region based on similarity to Homo sapiens; may contain additional exons not included here"
/codon_start=1
/product="general transcription factor I1H polypeptide 2"
/protein_id="NP82753.1"
/db_xref="GI:9082152"
/translation="MDPEPTKRWEGGYRTWEILKDETSLSKATIEDILFKARR
RYPEHGVRLMGHMLVYVDGSTMEDQDLKPNRLCTLLKLYFVEYFDQNPIS
QICIIYTKRAEKLTELSCNPRKHTLSKXAVDMTCHEPSLYNSLMQTKLEMP
GHTSREVLIIFFSLTTCDSNIYDLIKTLAKIRVSLGSLAEVAVCTVLARETGT
YHVLDETHYKELLAAHVSPASSSECSLIRMGFPQHTIASLSQDQAKSPSMAHL
DNNSTPGLTLGGYPCQCRKXCYCLPVECKICGLTVSAPHLARSVHFLPLDAQIE

Lamazares, R., Landers, T., Lehoczyk, J., Levine, R., Liu, G.,
 Maclean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M.,
 McEwan, P., McKernan, K., Murphy, T., Naylor, J., Nguyen, C., Norbu, C.,
 Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C.,
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S.,
 Severy, P., Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
 Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

Submitted (18-MAY-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collamore, A.,
 Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kanat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,
 Landers, T., Lehoczyk, J., Levine, R., Lindblad-Toh, K., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L.,
 Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

Submitted (14-APR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collamore, A.,
 Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,
 Landers, T., Lehoczyk, J., Levine, R., Lindblad-Toh, K., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L.,
 Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

Submitted (01-MAY-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On May 1, 2002 this sequence version replaced gi:20128009.

All repeats were identified using RepeatMasker:

Smit, A. F. A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center Project name: L10664

Center clone name: 47_G_4

FEATURES

source

Location/Qualifiers

1. 164314

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="18"

/map="18"

/clone="RP11-47G4"

/clone_lib="RPC1-11 Human Male BAC"

635..675

/rpt_family="(CA)n"

828..2491

/rpt_family="L2"

2507..2639

/rpt_family="Ricksha"

2650..2702

/rpt_family="Ricksha"

2696..3085

/rpt_family="Ricksha"

complement(3162..6085)

/rpt_family="L1PA7"

6086..6139

/rpt_family="(TG)n"

complement(6140..6987)

/rpt_family="L1PA7"

6988..7010

/rpt_family="(TTTA)n"

complement(7011..7047)

/rpt_family="L1PA7"

complement(7048..10232)

/rpt_family="L1PA3"

complement(10233..12805)

/rpt_family="L1PA7"

12811..13934

/rpt_family="Ricksha"

13938..14042

/rpt_family="Ricksha"

complement(14057..14132)

/rpt_family="MLTIC"

complement(14179..14679)

/rpt_family="MER67C"

14680..14953

/rpt_family="L1PB3"

complement(14954..15101)

/rpt_family="MER67C"

16096..16140

/rpt_family="AT rich"

complement(16290..16936)

/rpt_family="L1MD2"

complement(16958..19131)

/rpt_family="L1MB3"

complement(19136..19302)

/rpt_family="L1MD2"

20225..20249

/rpt_family="(A)n"

20890..21294

/rpt_family="MER57A"

21969..22082

/rpt_family="L2"

22209..22344

/rpt_family="MIR"

complement(22445..22737)

/rpt_family="L1PA8"

24066..24110

/rpt_family="A-rich"

```

repeat_region complement(24479..24697)
/rpt_family="L1ME3"
repeat_region complement(25011..25187)
/rpt_family="L1ME3"
repeat_region complement(25721..26013)
/rpt_family="AluSq"
repeat_region 26018..26039
/rpt_family="AT_rich"
repeat_region 27532..27695
/rpt_family="MIR"
repeat_region complement(27871..28207)
/rpt_family="AluJb"
repeat_region 28358..28378
/rpt_family="AT_rich"
repeat_region 28782..28870
/rpt_family="L1MED"
repeat_region 30177..30256
/rpt_family="MER91"
repeat_region 30740..30767
/rpt_family="AT_rich"
repeat_region 32285..32308
/rpt_family="AT_rich"
repeat_region complement(32338..32420)
/rpt_family="L1ME3"
repeat_region 32439..32666
/rpt_family="L1M4b"
repeat_region complement(32775..34045)
/rpt_family="L1MC3"
repeat_region complement(34114..34172)
/rpt_family="L1MC3"
repeat_region complement(34875..35263)
/rpt_family="L1PA1c"

```

Query Match 60.5%; Score 24.2; DB 9; Length 164314;

Best Local Similarity 78.4%; Pred.No. 2e+02;

Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

Qy      2  AAGTCAAAACATATGACTTACCGAATCGTGAAGT 38
Db      104579 AAGTCAAAACATATTTTAAAGTAATTTCTAAAT 104615

```

Search completed: May 26, 2004, 17:22:16

Job time : 560.41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 15:20:44 ; Search time 319.936 Seconds
(without alignments)
3115.905 Million cell updates/sec

Title: US-10-676-299-7
Perfect score: 23
Sequence: 1 ttaatcatatgcgttttggta 23

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sv.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	23	100.0	727	1	ECRPARSR	X16045 E. coli R-f
2	23	100.0	120826	1	AP005147	AP005147 Salmonell
3	19	82.6	176325	2	BX279525	BX279525 Danio rer
4	19	82.6	184543	5	BX571681	BX571681 Zebrafish
5	19	82.6	215891	5	AL928692	AL928692 Zebrafish
6	19	82.6	219601	2	BX649641	BX649641 Danio rer
7	19	82.6	253172	5	BX571883	BX571883 Zebrafish
8	18.8	81.7	620	3	AF129331	AF129331 Phranta
9	18.8	81.7	3727	6	AX713338	AX713338 Sequence
10	18.8	81.7	3727	9	AK055877	AK055877 Homo sapi
11	18.8	81.7	11447	1	AE011312	AE011312 Leptospir
12	18.8	81.7	39574	3	AF314193	AF314193 Drosophil
13	18.8	81.7	74881	2	AC020320	AC020320 Drosophil
14	18.8	81.7	112659	9	AC010677	AC010677 Homo sapi
15	18.8	81.7	119595	5	BX005186	BX005186 Zebrafish
16	18.8	81.7	137635	2	BX629344	BX629344 Danio rer
17	18.8	81.7	161411	2	AC102700	AC102700 Mus muscu
18	18.8	81.7	170869	3	AC011696	AC011696 Drosophil
19	18.8	81.7	171831	3	AC007473	AC007473 Drosophil
20	18.8	81.7	193708	9	AC005035	AC005035 Homo sapi
21	18.8	81.7	223489	5	BX248100	BX248100 Zebrafish
22	18.8	81.7	228048	2	AC122883	AC122883 Mus muscu
23	18.8	81.7	261690	3	AE003825	AE003825 Drosophil
24	18.8	81.7	269223	6	AR408762	AR408762 Sequence
25	18.8	81.7	269223	6	AX067466	AX067466 Sequence
26	18.8	81.7	295094	2	BX649429	BX649429 Danio rer
27	18.4	80.0	21479	3	CBRG03E09	CBRG03E09
28	18.4	80.0	34796	3	CEFS3C11	CEFS3C11
29	18.4	80.0	99431	8	AY224188	AY224188 Medicago
30	18.4	80.0	124457	8	AC146585	AC146585 Medicago
31	18.4	80.0	212669	2	AC111549	AC111549 Rattus no
32	18.4	80.0	261226	2	AC094788	AC094788 Rattus no
33	18.2	79.1	201	8	AJ598878	AJ598878 Arabidops
34	18.2	79.1	430	8	AY201116	AY201116 Arabidops
35	18.2	79.1	460	11	BX530054	BX530054 Arabidops
36	18.2	79.1	466	3	B0C506884	B0C506884 Buthus oc
37	18.2	79.1	466	3	B0C506893	B0C506893 Buthus oc
38	18.2	79.1	466	3	B0C506894	B0C506894 Buthus oc
39	18.2	79.1	466	3	B0C506896	B0C506896 Buthus oc
40	18.2	79.1	466	3	B0C506897	B0C506897 Buthus oc
41	18.2	79.1	466	3	B0C506901	B0C506901 Buthus oc
42	18.2	79.1	466	3	B0C506902	B0C506902 Buthus oc
43	18.2	79.1	486	11	BX530044	BX530044 Arabidops
44	18.2	79.1	491	3	MEU550701	MEU550701 Mesobuthu
45	18.2	79.1	491	3	MEU550702	MEU550702 Mesobuthu

ALIGNMENTS

RESULT 1	ECRPARSR	E. coli R-factor R773 arsr gene.	727 bp	DNA	linear	BCT 07-SEP-1994
LOCUS	ECRPARSR	E. coli R-factor R773 arsr gene.				
DEFINITION	X16045	GI:42716				
ACCESSION	X16045					
VERSION	X16045.1					
KEYWORDS	arsenical resistance; arsr gene; Arsr protein; DNA-binding protein; regulatory protein; resistance gene.					
SOURCE	Escherichia coli					
ORGANISM	Escherichia coli					
REFERENCE	1 (bases 1 to 727)					
AUTHORS	Rosen, B.P.					

Pred. No. is the number of results predicted by chance to have a

TITLE	Identification of the metalloregulatory element of the plasmid-encoded arsenical resistance operon
JOURNAL	Nucleic Acids Res. 18 (3), 619-624 (1990)
MEDLINE	90174986
PUBMED	2408017
REFERENCE	2 (bases 1 to 727)
AUTHORS	Rosen, B.P.
TITLE	Direct Submission
JOURNAL	Wayne State University, School of Medicine, 54- E Canfield Avenue, Detroit MI 48201, U S A
LOCATION	Location/Qualifiers
SOURCE	1..727
	/organism="Escherichia coli"
	/mol_type="genomic DNA"
	/db_xref="taxon:562"
	/clone="pWSU1"
	/clone_lib="pBR322"
promoter	73..79
	/note="pot. -35 region"
promoter	96..102
	/note="pot. -10 region"
misc_feature	107
	/note="transcriptional start site"
RBS	114..118
	/note="pot. ribosome binding site"
CDS	125..473
	/note="unnamed protein product; ArsR protein (AA 1 - 117)";
	/codon_start=1
	/transl_table=11
	/protein_id="CAA34168.1"
	/db_xref="GI:42717"
	/db_xref="GOA:P15905"
	/db_xref="SWISS-PROT:P15905"
	/translation="MQLTPQLFNLSDETRLGIVLLRMGBELCVCLMALDQSQ PKISRHLMVRESGILLDRKQKRWYRLSPHPWAQKIIEQAWLSQDDVQVIARK LASVNCSGSSKAVCI"
terminator	482..511
	/note="pot. stem-loop structure"
ORIGIN	
Query Match	100.0%; Score 23; DB 1; Length 727;
Best Local Similarity	100.0%; Pred. No. 36;
Matches	23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 TTAATCATATGCGTTTGGTTA 23
	42 TTAATCATATGCGTTTGGTTA 64
Db	
RESULT 2	
AP005147	120826 bp DNA circular BCT 14-MAY-2002
LOCUS	Salmonella typhimurium plasmid R64 DNA, complete sequence.
DEFINITION	
ACCESSION	AP005147.1 GI:20521502
VERSION	
KEYWORDS	Salmonella typhimurium
SOURCE	Salmonella typhimurium
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE	1
AUTHORS	Komano, T., Kubo, A. and Nishioka, T.
TITLE	Shufflon: multi-inversion of four contiguous DNA segments of plasmid R64 creates seven different open reading frames
JOURNAL	Nucleic Acids Res. 15 (3), 1165-1172 (1987)
MEDLINE	87146423
PUBMED	3023698
REFERENCE	2
AUTHORS	Kubo, A., Kusukawa, A. and Komano, T.
TITLE	Nucleotide sequence of the rci gene encoding shufflon-specific DNA recombinase in the IncII plasmid R64: homology to the site-specific recombinases of interease family

[illegible]

Mol. Gen. Genet. 213 (1), 30-35 (1988)

89127142
3065610

3
Komano, T., Toyoshima, A., Morita, K. and Nisioka, T.
Cloning and nucleotide sequence of the *oriT* region of the IncII plasmid R64
J. Bacteriol. 170 (9), 4385-4387 (1988)

88314948
3045094

4
Furuya, N., Nisioka, T. and Komano, T.
Nucleotide sequence and functions of the *oriT* operon in IncII plasmid R64
J. Bacteriol. 173 (7), 2231-2237 (1991)

91177811
1848841

5
Furuya, N. and Komano, T.
Determination of the nick site at *oriT* of IncII plasmid R64: global similarity of *oriT* structures of IncII and IncP plasmids
J. Bacteriol. 173 (20), 6612-6617 (1991)

92011438
1917882

6
Kim, S.R., Funayama, N. and Komano, T.
Nucleotide sequence and characterization of the *traABCD* region of IncII plasmid R64
J. Bacteriol. 175 (16), 5035-5042 (1993)

93352408
8349545

7
Furuya, N. and Komano, T.
Surface exclusion gene of IncII plasmid R64: nucleotide sequence and analysis of deletion mutants
Plasmid 32 (1), 80-84 (1994)

95083745
7891676

8
Furuya, N. and Komano, T.
Nucleotide sequence and characterization of the *trbABC* region of the IncII plasmid R64: existence of the *pnd* gene for plasmid maintenance within the transfer region
J. Bacteriol. 178 (6), 1491-1497 (1996)

96198148
8626273

9
Kim, S.R. and Komano, T.
The plasmid R64 thin pilus identified as a type IV pilus
J. Bacteriol. 179 (11), 3594-3603 (1997)

97315231
9471405

10
Narahara, K., Rahman, E., Furuya, N. and Komano, T.
Requirement of a limited segment of the *sog* gene for plasmid R64 conjugation
Plasmid 38 (1), 1-11 (1997)

97428559
9281491

11
Furuya, N. and Komano, T.
Mutational analysis of the R64 *oriT* region: requirement for precise location of the *Nika*-binding sequence
J. Bacteriol. 179 (23), 7291-7297 (1997)

98053841
9393692

12
Yoshida, T., Furuya, N., Ishikura, M., Isobe, T., Haino-Fukushima, K., Ogawa, T. and Komano, T.
Purification and characterization of thin pili of IncII plasmids ColiB-p9 and R64: formation of PilV-specific cell aggregates by type IV pili
J. Bacteriol. 180 (11), 2842-2848 (1998)

```

MEDLINE      98268996
PUBMED      9603870
REFERENCE   13
AUTHORS     Komano,T., Yoshida,T., Narahara,K. and Furuya,N.
TITLE       The transfer region of IncII plasmid R64: similarities between R64
JOURNAL     Mol. Microbiol. 35 (6), 1348-1359 (2000)
MEDLINE     20233621
PUBMED     10760136
REFERENCE   14
AUTHORS     Sampei,G., Komano,T., Sasaki,T., Tachibana,K., Furuya,N., Saito,Y.,
TITLE       Suzuki,T. and Mizobuchi,K.
JOURNAL     Organization and diversification of plasmid genomes: complete
AUTHORS     nucleotide sequence of R64 genome
TITLE       15 (bases 1 to 120826)
JOURNAL     Komano,T., Sasaki,T., Tachibana,K., Furuya,N., Saito,Y., Suzuki,T.,
AUTHORS     Direct Submission
TITLE       Submitted (07-MAY-2002) Gen-ichi Sampei, The University of
JOURNAL     Electro-Communications, Applied Physics and Chemistry; Chofugaoka
AUTHORS     1-5-1, Chofu, Tokyo 182-8585, Japan (E-mail:sampeigpc.uec.ac.jp,
TITLE       Tel:81-424-43-5481, Fax:81-424-43-5501)
JOURNAL     Location/Qualifiers
FEATURES
source
    /organism="Salmonella typhimurium"
    /mol_type="genomic DNA"
    /db_xref="taxon:602"
    /lab_host="Escherichia coli strain K-12"
    /plasmid="R64"
    /note="R64 strain drd-11"
    complement(join(20047..20418,3609..4147))
    /notes="100 pct identical to sp:Y122_ECOLI[hypothetical
    protein of insertion sequence IS2]
    location complement(4143..4147) and location
    complement(20047..20051) are duplicated"
    378..467
    /gene="repY"
    378..467
    /genes="repY"
    /notes="100 pct identical to pir:A35445[RepY of plasmid
    Colib-P9]
    /codon_start=1
    /transl_table=11
    /protein_id="BAB91567.1"
    /db_xref="GI:20521503"
    /translation="MKPYQRFNPVQCINRHRNSAISDSLQV"
    455..1486
    /genes="repZ"
    455..1486
    /gene="repZ"
    /notes="100 pct identical to sp:REP2_ECOLI.sp:P18023[RepZ
    of plasmid Colib-P9]"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB91568.1"
    /db_xref="GI:20521504"
    /translation="MAGLKNTPYNAVHWSQLAPEQIRFWEDYEGRATTFUVEPERK
    RYKRRGESHSTKPKCSNWSYPERKALKQGLHAYNLVKDQPVTEQSLRMRSR
    HPFYVKRTFVGKAYRPERKQSLDAIWPVLVSFSDACTHTVGMSTVLAEEISPKD
    SEGHVPLEVTVSRLLRLAEQVRFGLVSEETWDRHQRLLPRYVWIPAGQM
    LGVDMVKLHQQQKRLRESIRQOLIREGLVREDEDI SVHAAKRWYLORSODALKR
    REKAASKRANRLKKLPVQIYEMAEYLRKRLPPDEAYFCSDDHLKRLAIRLQLE
    LTLAAPPH"
    2761..3114
    /gene="arsR"
    2761..3114
    /genes="arsR"
    /notes="100 pct identical to sp:ARR1_ECOLI.sp:P15905[arsR
    of plasmid R773]"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB91569.1"
    /db_xref="GI:20521505"
    /translation="MLQITPLQLFKNLSDETRIGIVLLIREMGLCVCDLQMALDQSQ
    PRISRELAHMLRESGILLDRKQKWHYRLSPHPWAAQIIHQAMLQQDDVQVITARK
    LASVNCSSSKAVCI"
    3162..3524
    /gene="arsD"
    3162..3524
    /gene="arsD"
    /note="99 pct identical to sp:ARD1_ECOLI.sp:P46003[ArSD of
    plasmid R773]"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB91570.1"
    /db_xref="GI:20521506"
    /translation="MKTLMVDFPAMCCSTGVTVDQALVDFSDVQVWLKQCGVOIR
    RFNLAQPMFSQVNEKVKAPTEASGARGELPLLLDGETVMAGRYKRAELARWFGIPL
    DKVGLAPSGCCCGCTSCC"
    join(3542..3597,20833..22533)
    /gene="arsA"
    join(3542..3597,20833..22533)
    /gene="arsA"
    /note="99 pct identical to pir:A25937[ArsA of plasmid
    R773, arsenical pump-driving ATPase]
    location 3593..3597 and location 20833..20837 are
    duplicated"
    join(3598..4147,20047..20832)
    /insertion_seq="IS2"
    complement(join(3609..4147,20047..20418))
    /gene="yadA"
    join(4148..7818,18208..20046)
    /note="different from TnS393[gb:M96392] in IS1133 insert
    position and terminal direct repeat"
    Query Match 100.0%; Score 23; DB 1; Length 120826;
    Best Local Similarity 100.0%; Pred. No. 8.3; Indels 0; Gaps 0;
    Matches 23; Conservative 0; Mismatches 0;
    QY 1 TTAATCATATGCGTTTGGTTA 23
    DB 2678 TTAATCATATGCGTTTGGTTA 2700
    RESULT 3
    BXZ79525/c
    LOCUS
    DEFINITION
    BXZ79525
    VERSION
    BXZ79525.3 GI:37719061
    KEYWORDS
    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
    SOURCE
    Danio rerio (zebrafish)
    ORGANISM
    Danio rerio
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
    Cypriniformes; Cyprinidae; Danio.
    1 (bases 1 to 176325)
    McIaren,S.
    Direct Submission
    Submitted (17-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,
    Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
    zfish-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
    On Oct 18, 2003 this sequence version replaced gi:35208834.
    COMMENT
    ----- Genome Center
    Center: Wellcome Trust Sanger Institute
    Center code: SC
    Web site: http://www.sanger.ac.uk
    Contact: zfish-help@sanger.ac.uk
    ----- Project Information
    Center project name: zC244C18
    ----- Summary Statistics
    Assembly program: XCAP4; version 4.5
    Chemistry: Dye-terminator; 100% of reads

```

Consensus quality: 175000 bases at least Q40
 Consensus quality: 175423 bases at least Q30
 Consensus quality: 175704 bases at least Q20
 Insert size: 185227; 4.4% error; agarose-fp
 Quality coverage: 10.83x in Q20 bases; sum-of-contigs Quality
 coverage: 10.36x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 68108: contig of 68108 bp in length
 * 68208: gap of 100 bp
 * 68209: 72158: contig of 3950 bp in length
 * 72159 72258: gap of 100 bp
 * 72259 158147: contig of 85889 bp in length
 * 72259 158147: contig of 85889 bp in length
 * 158148 158247: gap of 100 bp
 * 158248 176325: contig of 18078 bp in length.

FEATURES
 source
 Location/Qualifiers
 1..176325

/organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="CH211-244C18"
 /clone_lib="CHORI-211"

misc_feature
 1..68108
 /note="assembly_fragment:01778
 clone_end:T7
 vector_side:left"

misc_feature
 68209..72158
 /note="assembly_fragment:00514"

misc_feature
 72259..158147
 /note="assembly_fragment:00910"

misc_feature
 158248..176325
 /note="assembly_fragment:01439
 clone_end:SP6
 vector_side:right"

ORIGIN

Query Match 82.6%; Score 19; DB 2; Length 176325;
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TTAATCATATGCGTTTGTG 19

Db 152847 TTAATCATATGCGTTTGTG 152829

RESULT 4

LOCUS BX571681 184543 bp DNA linear VRT 06-NOV-2003
 DEFINITION Zebrafish DNA sequence from clone DKEY-73N10 in linkage group 18,
 complete sequence.

ACCESSION BX571681.7 GI:38141682

VERSION BX571681

KEYWORDS HTG.

SOURCE Danio rerio (zebrafish)

ORGANISM

Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 184543)

Phillimore, B.

Direct Submission

Submitted (06-NOV-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

zfsh-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Oct 31, 2003 this sequence version replaced gi:37936371.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: zfsh-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest, except on the rare
 occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
 on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
 Zebrafish PUC subclones occasionally display inconsistency over the
 length of mononucleotide A/T runs and conserved TA repeats. Where
 this is found the longest good quality representation will be
 submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat
 discovery system (Zhiyong Bao and Sean Eddy, submitted), and those
 beginning 'dr' were identified by Rick Waterman (Stephen Johnson
 lab, WashU). For further information see

http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-73N10
 is from a Zebrafish BAC library

VECTOR: pIndigoBAC-5.

FEATURES
 source

Location/Qualifiers
 1..184543
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="DKEY-73N10"
 /clone_lib="DanioKey"

ORIGIN

Query Match 82.6%; Score 19; DB 5; Length 184543;
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TTAATCATATGCGTTTGTG 19

Db 156354 TTAATCATATGCGTTTGTG 156372

RESULT 5

LOCUS AL928692 215891 bp DNA linear VRT 20-NOV-2003
 DEFINITION Zebrafish DNA sequence from clone CH211-208D15 in linkage group 12,
 complete sequence.

ACCESSION AL928692

VERSION AL928692.9 GI:38488525

KEYWORDS HTG.

SOURCE Danio rerio (zebrafish)

ORGANISM

Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 215891)

Leongamornlert, D.

Direct Submission

Submitted (20-NOV-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

zfsh-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk

COMMENT On Nov 20, 2003 this sequence version replaced gi:31043727.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep

Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhiyong Bao and Sean Eddy, submitted), and those beginning 'dr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml

CH211-208D15 is from a CHORI-211 BAC library
 VECTOR: pTRAC2.1.
 Location/Qualifiers
 1..215891
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="CH211-208D15"
 /clone_lib="CHORI-211"

ORIGIN
 Query Match 82.6%; Score 19; DB 5; Length 215891;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTAATCATATCGGTTTTG 19
 |||||
 Db 145868 TTAATCATATCGGTTTTG 145850

RESULT 6
 BX649641/c
 LOCUS
 DEFINITION Danio rerio clone CH211-22301, WORKING DRAFT SEQUENCE, 3 unordered pieces.
 ACCESSION BX649641
 VERSION BX649641.7 GI:38198314
 KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 219601)
 Phillimore,B.
 Direct Submission
 REFERENCE
 AUTHORS Submitted (04-NOV-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Nov 7, 2003 this sequence version replaced gi:38143432.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: zfish-help@sanger.ac.uk

Center project name: zc22301
 ----- Project Information
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 219100 bases at least Q40
 Consensus quality: 219178 bases at least Q30
 Consensus quality: 219278 bases at least Q20
 Insert size: 219401; sum-of-contigs
 Insert size: 217356; 3.8% error; agarose-fp
 Quality coverage: 8.13x in Q20 bases; sum-of-contigs Quality coverage: 8.37x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 50258: contig of 50258 bp in length
 50359: gap of 100 bp
 50359 151175: contig of 100817 bp in length
 151176 151275: gap of 100 bp
 151276 219601: contig of 68326 bp in length.

FEATURES
 source
 1..219601
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="CH211-22301"
 /clone_lib="CHORI-211"
 1..50258
 /note="assembly_fragment:00500"
 50359..151175
 /note="assembly_fragment:01154"
 151276..219601
 /note="assembly_fragment:01977"

ORIGIN
 Query Match 82.6%; Score 19; DB 2; Length 219601;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTAATCATATCGGTTTTG 19
 |||||
 Db 45832 TTAATCATATCGGTTTTG 45814

RESULT 7
 BX571883
 LOCUS
 DEFINITION Zebrafish DNA sequence from clone DKRY-151P17 in linkage group 18, complete sequence.
 ACCESSION BX571883
 VERSION BX571883.3 GI:34956448
 KEYWORDS HTG.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 253172)
 Wood,J.
 Direct Submission
 REFERENCE
 AUTHORS
 TITLE

JOURNAL Submitted (23-SEP-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clon requests: clonerequest@sanger.ac.uk
 COMMENT On Sep 23, 2003 this sequence version replaced gi:33386634.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhikang Bao and Sean Eddy, submitted), and those beginning 'dr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see
http://www.sanger.ac.uk/Projects/D_reio/fishmask.shtml
 DKEY-151P17 is from a Zebrafish BAC library
 VECTOR: pIndigoBAC-5.

FEATURES
 source
 1..253172
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone_lib="DKEY-151P17"
 /clone_lib="DanioKey"

ORIGIN

Query Match 82.6%; Score 19; DB 5; Length 253172;
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTATCATATCGGTTTGG 19

Db 215479 TTATCATATCGGTTTGG 215497

RESULT 8
 AF129331 620 bp DNA linear INV 18-MAY-2001
 LOCUS Phrantela marginata cytochrome c oxidase subunit I gene, partial cds; mitochondrial gene for mitochondrial product.
 DEFINITION
 ACCESSION AF129331
 VERSION AF129331.1 GI:5081538

KEYWORDS mitochondrion Phrantela marginata
 SOURCE Phrantela marginata
 ORGANISM Eukaryota; Metazoa; Mollusca; Gastropoda; Orthogastropoda; Apogastropoda; Caenogastropoda; Sorbeoconcha; Hypsogastropoda; Littorinimorpha; Rissoidae; Hydrobiidae; Phrantela.
 REFERENCE
 1 (bases 1 to 620)
 AUTHORS Hershtler,R., Liu,H.P. and Mulvey,M.

TITLE Phylogenetic relationships within the aquatic snail genus Tryonia: implications for biogeography of the North American Southwest
 JOURNAL Mol. Phylogenet. Evol. 13 (2), 377-391 (1999)
 MEDLINE 20072936
 PUBMED 10603265
 REFERENCE 2 (bases 1 to 620)
 AUTHORS Liu,H.P.
 DIRECT SUBMISSION
 TITLE Direct Submission
 JOURNAL Submitted (18-FEB-1999) Biology, Southwest Missouri State University, 901 S. National Ave., Springfield, MO 65804, USA
 FEATURES
 Location/Qualifiers
 source
 1..620
 /organism="Phrantela marginata"
 /organelle="mitochondrion"
 /mol_type="genomic DNA"
 /db_xref="taxon:93097"
 /country="Australia:tributary of Thirteen Mile Creek"
 <1..>620
 /codon_start=1
 /transl_table=5
 /product="cytochrome c oxidase subunit I"
 /protein_id="AAD39432.1"
 /db_xref="GI:5081539"
 /translation="SGVVGTAISLLIRAEIQGQALLGDDOLYNNIVTAHAFVMIFFLLVPMHIGGFGNVLVPLMGADPMAPFNNMFWLLPALLLSAAVSGAGTGTWYVPLASNLHAGSVDIAIFSLHAGVSSILGAVNFTTINRWGMQFRLPLFVWSVKITAILLLSLPVLAVGAVTMLLTDRTNFTAFDPAGGDPILY"

ORIGIN

Query Match 81.7%; Score 18.8; DB 3; Length 620;
 Best Local Similarity 90.9%; Pred. No. 1.5e+03;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTATCATATCGGTTTGGTT 22

Db 215 TTATCATATCGGTTTGGTT 236

RESULT 9
 AX713938/c 3727 bp DNA linear PAT 15-APR-2003
 LOCUS AX713938
 DEFINITION Sequence 622 from Patent EP1293569.
 ACCESSION AX713938
 VERSION AX713938.1 GI:29888866
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1
 AUTHORS Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,O.I., Isono,Y., Hio,Y., Otsuka,K., Negai,K., Irie,R., Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Negahari,K. and Masuko,Y.
 Full-length cDNAs
 Patent: EP 1293569-A 622 19-MAR-2003;
 JOURNAL Helix Research Institute (JP) ; Research Association for Biotechnology (JP)
 FEATURES
 Location/Qualifiers
 source
 1..3727
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Query Match 81.7%; Score 18.8; DB 6; Length 3727;
 Best Local Similarity 90.9%; Pred. No. 1e+03;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTATCATATCGGTTTGGTT 22

Db 2104 TTATCATATCGGTTTGGTT 2083

RESULT 10	AK055877	3727 bp	mrna	linear	PRI 01-AUG-2002	
LOCUS	AK055877/c					
DEFINITION	homo sapiens cDNA FLJ31315 fis, clone LIVER1000303.					
ACCESSION	AK055877					
VERSION	AK055877.1	GI:16550712				
KEYWORDS	cdigo capping; fis (full insert sequence).					
SOURCE	homo sapiens (human)					
ORGANISM	homo sapiens					
REFERENCE	1					
AUTHORS	Taahiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukumori, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, I., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yasashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, E., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugeno, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.					
TITLE	NEDO human cDNA sequencing project					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 3727)					
AUTHORS	Isogai, T., Otsuki, T. and Sugiyama, T.					
TITLE	Direct Submission					
JOURNAL	Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)					
COMMENT	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.					
FEATURES	Location/Qualifiers					
source	1..3727					
	/organism="Homo sapiens"					
	/mol_type="mrna"					
	/db_xref="taxon:9606"					
	/clone="LIVER1000303"					
	/tissue_type="liver"					
	/clone_lib="LIVER1"					
	/note="Cloning vector: pME18SFL3"					
ORIGIN						
Query Match	81.7%;	Score 18.8;	DB 9;	Length 3727;		
Best Local Similarity	90.9%;	Pred. No. 1e+03;				
Matches	20;	Conservative	0;	Mismatches	2;	
			Indels	0;	Gaps	0;
QY	1	TTAATCATATCGCTTTTGTT	22			
Db	2104	TTAATCATATCGCTTTTGTT	2083			
RESULT 11	AE011312	11447 bp	DNA	linear	BCT 28-APR-2003	
LOCUS	AE011312					
DEFINITION	Leptospira interrogans serovar lai str. 56601 chromosome I, section 121 of 397 of the complete sequence.					
ACCESSION	AE011312					
VERSION	AE011312.1	GI:24194885				
KEYWORDS	Leptospira interrogans serovar lai str. 56601					
SOURCE	Leptospira interrogans serovar lai str. 56601					
ORGANISM	Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.					
REFERENCE	1 (bases 1 to 11447)					
AUTHORS	Ren, S.X., Fu, G., Jiang, X.G., Zeng, R., Miao, Y.G., Xu, H., Zhang, Y.X., Xiong, H., Lu, F., Lu, F., Jiang, H.Q., Jia, J., Tu, Y.F., Jiang, Y.X., Gu, W.Y., Zhang, Y., Cai, Z., Sheng, H.H., Yin, H.F., Zhang, Y., Zhu, G.F., Wan, M., Huang, H.L., Qian, Z., Wang, S.Y., Ma, W., Yao, Z.J.,					
TITLE	Unique physiological and pathogenic features of Leptospira interrogans revealed by whole-genome sequencing					
JOURNAL	Nature 422 (5934), 888-893 (2003)					
PMED	12712204					
REFERENCE	2 (bases 1 to 11447)					
AUTHORS	Ren, S., Fu, G., Jiang, X., Zeng, R., Xiong, H., Lu, L., Lu, G., Jiang, H., Ding, Y., Jia, J., Tu, Y., Gu, W., Cai, Z., Sheng, H., Yin, H., Zhang, Y., Zhu, G., Wang, S., Shen, Y., Qiang, B., Chen, Z., Wen, Y., Xu, J. and Zhao, G.					
TITLE	Direct Submission					
JOURNAL	Submitted (12-MAR-2002) Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Shanghai 201203, China					
COMMENT	Updated information will be available at our World Wide Web site (http://www.chgc.sh.cn/lep/). Comments to the authors are appreciated.					
FEATURES	Location/Qualifiers					
source	1..11447					
	/organism="Leptospira interrogans serovar lai str. 56601"					
	/mol_type="genomic DNA"					
	/strain="56601"					
	/serovar="Lai"					
	/db_xref="taxon:189518"					
	/chromosome="I"					
	/notes="Strain 56601 is maintained by the Institute of Epidemiology and Microbiology, Chinese Academy of Preventive Medicine"					
gene	144..701					
	/locus_tag="LA1317"					
CDS	144..701					
	/locus_tag="LA1317"					
	/notes="confirmed by proteomics"					
	/codon_start=1					
	/transl_table=11					
	/product="unknown protein"					
	/protein_id="AA048516.1"					
	/db_xref="GI:24194886"					
	/translation="MFCSSNMLTHFGQAGKGPFCILILFLSLAIFPIEQMKETPK SKDFVLIVLSKQDFNDEKIKAVIRFNVVSQQLSYGFDGLKDFNYILECYALKK KNGGPPQALRTQHGEMWAKSQNEETMGALNSGKELVNTISVSRQCLTMTGTGYQFV KHRYPISKKEIIVSDPIVRVHE"					
gene	841..981					
	/locus_tag="LA1318"					
CDS	841..981					
	/locus_tag="LA1318"					
	/codon_start=1					
	/transl_table=11					
	/product="hypothetical protein"					
	/protein_id="AA048517.1"					
	/db_x					

```

/transi_table=11
/product="hypothetical protein"
/protein_id="AA048519.1"
/db_xref="GI:24194889"
/translation="MGYRAFCSLGFRLKKEPFEIGS3SKLTSSBITLGMELTSGFLI
EA"
gene
2804..4213
/gene="dsbD"
/locus_tag="LA1321"
CDS
2804..4213
/gene="dsbD"
/locus_tag="LA1321"
/codon_start=1
/transi_table=11
/product="thiol:disulfide interchange protein DsbD"
/protein_id="AA048520.1"
/db_xref="GI:24194890"
/translation="WKHSRNOISPLVFLVLSKLIYNNELIFPSILSILGCVLP
ISAVSSSLRKIVVEGSGDPTGLIWMGGLASVILRPIVNLIFITVIGILRARGE
GPKILHEAIYILGVVYISCFGLIWSFGAGFNVILRPIVNLIFITVIGILRARGE
DLIHPPFRSEKVTGQCGGTFLLMGAGLLSPCGVGVVAILLQITAGAVISIG
SVFLASPMFLFGVGLFPLMIGVGLSPKSKRWEMWQIWLICFVLYSYTYFQK
SLSGVSPDSIPLAAGFILFLFCLYFLPDENRYVVKMKTLILGGIILASVGLV
LLMNLNETNISIVVEMETKGNLSWHRIPSKAYAGTGTGKKVIFIDFYADWCTCKTF
BELTOSDALANDALOGALLKIQDDPIFEIYAKDPREELKIGLFPFVILDAGNLL
YNTDYQDTSTNIQLKQL"
4973..5080
/locus_tag="LA1322"
CDS
4973..5080
/locus_tag="LA1322"
/codon_start=1
/transi_table=11
/product="hypothetical protein"
/protein_id="AA048521.1"
/db_xref="GI:24194891"
/translation="WIALQVIANHLFLIRYTKLVLFRFYEDFTTRTYE"
5570..7807
/gene="purl"
/locus_tag="LA1323"
CDS
5570..7807
/gene="purl"
/locus_tag="LA1323"
/EC_number="6.3.5.3"
/note="FGAM synthase II"
/codon_start=1
/transi_table=11
/product="phosphoribosylformylglycinamide synthase II"
/protein_id="AA048522.1"
/db_xref="GI:24194892"
/translation="MEKDVSLDALBEGLTKBEFKQIEILGRIPNSTELGFSAMW
SEHCYSKNSILKLTPTTSDKLAKAGEENAGAMDIGDLAVVFKIESNHPITAVFP
YQGAATGGVIMRDIPTMGARPIVLSLRFNGNDPEPKYLLSRVKGIDYGNLSG
IAVSGELPIIDBCTSKNPLNVAVTGIVRHQDASATTGGQIGNSVIVGATTCRDGI
HGASPASDLKSKESKSAVQVGDPMFKLLMEASLEAIQKGLLVGIQDMGAIGIS
ATSEMSKGTGMKLDLDLVPFRITGNAYEAMLSSEQRMVLPVFKKSELSVIFE
KLNNAVYKIGSVTADGMIEHITWGGKLGAKIPAESLVLGCGAPRYETKAPSYLDVAK
TWKDEPIDVTGANSKEILLKILSSMNVCSRKPITEQYDSEVLGKLIIGPLDGGVS
AIPGNKALATPCNSRYTYLDYPKGAFAVCAEARNVYVGTATPYGNNLNFANP
YIPENYTFSCIRGMDACRFLGTPVGTGNVSPKNSPEGPITPTTIGVGLLENK
KLKLPNFKIGVELIADNPNRPSLGGSEYLVKIHGOINGSPELDIKLEELCKLIL
SLNRSILKSKDLSIAGIAVALSKTVLFSGLGIESDLTSLRRRLDLTLFGESSTAV
LVGFDLSLKEIDIRQTEAYGLKFPYICKNTSSSGILEIKDAEIKSPQELSGPYEKGLE
NYPAL"
7823..8476
/locus_tag="LA1324"
CDS
7823..8476
/locus_tag="LA1324"
/codon_start=1
/transi_table=11
/product="leucine-rich-repeat containing protein"
/protein_id="AA048523.1"
/db_xref="GI:24194893"
/translation="MNFSLKKIFKLYFLVWVSFGIDCKKNAVEILEEAKRSSESIG

```

```

ILDLGMOKLTSIPBGVCSFENLTQDLRLNSLSLPCWIGACKNLEQINLFGNDLNTV
PSSPSKLNKLVLLIGNNDFTPLPSSELLFLPLKTLTLDQNKLTLPETDVEILLASLS
LESLDLNLSGKIVPSNYNKLKNIHLKLNLIKTSKGDADAKLQVLPNTRIDI"
8484..11228
/gene="ileS"
/locus_tag="LA1325"
CDS
8484..11228
/gene="ileS"
/locus_tag="LA1325"
/EC_number="6.1.1.5"
/codon_start=1
/transi_table=11
/product="isoleucyl-tRNA synthetase"
/protein_id="AA048524.1"
/db_xref="GI:24194894"
/translation="MSETKENPYSTVLLPKTDFPMKADLKEEPAQIQSWKSGQIF
RKMKQRKKEKFEVLHGGPYANGNPHLSINKIVKDIKSKSLAGFYADMLPGWD
CHGLPEVQVNLKMLKKARETGEEELRLRCNVAEAPVKGQGLDLSFLCFWEEDRIY
KTMSPGEAKIVFVGFELFQNGVYIRGKPVYWSIDLATAHAAEIEIYFHVSPSIVY
KPTIVGEKQKFLIWTITTPANLIGICFNRIEYFFKFTSEGEELILADGLAESVT
DTYGLNLYKYSISSELSVLKQHPFMDRISISLFDHVTLDAGTCVHTAFHGQD
STVGLAAGLEPSPVDVGYKTDDEPLMGQKRVFANDPEIIQLKDKGILLHSEPE
HSVPHWSRSKKPLIFRATPOWPFKMDPNELREKLSAIDGVOMIWSMGITRIRSMVET
RPMKLSORNMVGPVIPAFTCESCGOITHIDVSIQFTKVRKGLIETWYSEKAADLL
PPTCKKCKSDSPFKGNDILDWFFSGVSSPAVLGERKDEPPADLYFGSDQHRGHP
QSSLWSPMALRGIPPKYKTLVTHGVLDKRGHPSKNSIGINDIDTADVIQIYGADILKL
WVSSLDPRDDIKVKGESLKIIVSEYKIRINTFRYLLNGLDGHTSEQNLPFELEBBL
FKSKLSQFAEDVIANYEAYQHQIYQKGLFCTVTLTSLQDFMDIRDMYCDARNSKT
RKSATALQVILETLCITAPILSFSTAEVWSNGKDSVFLFPDLPKRWKQNLEE
KFBSVLQVREVQKALEIKARQENKGLKLEAGLEIVSKNGTNTLTKILPKETLILFVV
"
SOVHEKNPGLDVLSEYENKAFSVKVLKPVQVCEPCRWHTEDISKEDLCGRCKSVVG
"
ORIGIN
Query Match 81.7%; Score 18.8; DB 1; Length 11447;
Best local Similarity 90.9%; Pred. No. 7.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 TTAATCATATGCGTTTTCGTT 22
Db 3518 TTAATGATAGCGTTTTCGTT 3539
RESULT 12
AF314193 LOCUS 39574 bp DNA linear INV 11-APR-2003
DEFINITION Drosophila melanogaster Toutatis (tou) gene, complete cds.
ACCESSION AF314193
VERSION AF314193.1 GI:12642597
KEYWORDS Drosophila melanogaster (fruit fly)
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 39574)
AUTHORS Fauvarque,M.O., Laurenti,P., Boivin,A., Bloyer,S., Griffin-Shea,R.,
Bourbon,H.M. and Dura,J.M.
TITLE Dominant modifiers of the polyhomeotic extra-sex-combs phenotype
induced by marked P element insertional mutagenesis in Drosophila
Genet. Res. 78 (2), 137-148 (2001)
JOURNAL 21589546
MEDLINE 11732091
REFERENCE 2 (bases 1 to 39574)
AUTHORS Bourbon,H.M.
TITLE The novel trx-G gene toutatis encodes a Drosophila member of the
WAL family of chromatin-associated proteins
Unpublished
JOURNAL 3 (bases 1 to 39574)
REFERENCE Bourbon,H.M.
AUTHORS Direct Submission
TITLE Submitted (18-OCT-2000) Centre de Biologie du Developpement,
JOURNAL CNRS/UPS, 118 Route de Narbonne, Toulouse, HG 31062, France

```


FEATURES
source

```

Location/Qualifiers
1. .39574
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="2"
/map="48A3-6"
1651. 38602
/genes="tou"
join(1651. 1903,25307. 25489,25552. 25797,26308. 26795,26857. 27479,
26857. 27479,27542. 28784,28847. 29119,29355. 29711,30457.
34467,35337. 35854,36084. 37248,37313. 37434,37500.
38602)
/genes="tou"
/product="Toutatis"
join(25363. 25489,25552. 25797,26308. 26795,26857. 27479,
27542. 28784,28847. 29119,29355. 29711,30457. 34467,
35337. 35854,36084. 37248,37313. 37434,37500. 37556)
/genes="tou"
/notes="putative subunit of a chromatin-remodeling complex"
/codon_start=1
/product="Toutatis"
/protein_id="AAK00302.1"
/db_xref="GI:12642598"
/translation="MNKAGDGDGKSNKSNAGGGGAGGPHDPTGLLDASLPAY
KGRDPTGAAAAASNLFNLSQFNAAAAGLGLPQAGGASANDRYSMAAAAAAGAH
HGNTHVAASQASLAGLHPASWMAQADQYFSLQASGLSPFPHPDLAAAFGP
AGMGGGGAGAGAGAGGAGVGLGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
QQQQQQOOLANSINAAAAAANPAAALASMHGFGVPGTSTIPSVSTGSGSI
STNSGGGKYSKSTASAYKPGTSTSSMANNPGSAVDPTVTHKELLAMQAAASGS
GSGSGGKSGSGSSSSSHSISGMISGGGGGMMWSGKASTSVASNSISLGVHPSLT
SNPLMSPHAGGSSSSSSSKDQKNPSLNALNSISQFQALGMPQQSQMAANAPA
ASTGVPSPATVSSPHSSQQQQQXGKSGSTSGSGSSKSTYMGTSHPHSLGVR
PPTETIITTSVIGPKIPGTTSRGRKTTISETOQTTOQQQHQHAEHLLOQQQ
AQKELDSITNA-SLLAFPLCPSPAKRLEMEYAAAMAAHQHQHQHQHQHQHQ
GMLGAGIPCMAGLVGLPQWGNFPLQSLVSXASSTAPTSTSSAGSNLQNSNS
DRVEVILPTTIDTNGAYLNSKKEVHDLTTDMATNSGCVNLSLKNAGSSALTPSG
AVGSASNPITIDEDAPLNSMKPSKSSSSSSSSSSSSSSSSSSSSSSSSSSSS
SGSGNSLOSLSITIAAGGTGCPGSGISGSGGTSPAGAGSGATGCGSGSGSGG
GSSYKGRFRLGRGVSPKNTVSLAQAGVGLKAPLQATQQLQGLGRTVIEKRL
ALSEANAHETSTDSVAASGLSESEDANILNVALVPELGLGWRTGRTVIEKRL
KQGTGRTVYAPGTTPLKNGQVFALIEQPSNLRENFSFARAIIVGFLQAP
PPYANDGEYIRMTDEDVAKRLDKVFTQTLNVEQRIEIAKQAMRDAKLQXEL
ARNKEKARQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK
KHQVLDRLLRERRMAERKDAEILQILRRPNEDEMPQELVPELQIAGNLPQ
AMADLLMVEFLHNFETIGPMESLPSLQNLHDLMSNADAEHELSSVTHLLVC
ALSDGVPNFPGRHTLLGSLRNADITNSVSEILRIYIATATGEVFMQGVLDRE
RRRVPDHLQSDTTTHSHSVKQNYKLLHNDTWKLSQKDRPFPVLPNTRAQ
MLAHLCDLNMKAVRLQIDGSLCTCAQMRKEKTYMTQKRYKALHKEKRIEAYR
AOAEREAAMQALMAQKLDARLKAEEEAASAAAEEAAAGTDGEATKGSPPNEKP
EDQDNEBGAKEPQQQQQPMVEGVDEASLVSFAKTI IOTDNLSTPKGSDMPTPT
YOLNGSSTPTTSVTGDMNVLLQAKKSGARNSINDEHHDVSI IIDDLSLDSLTN
VSEDEDNRLSABELOKLDKI VRASLNCKEALKSTNQLRAACFGQDRFWRYKWLKP
ACGIFIEALASQNDICDTHEALEAMDDKQANDKENSENEKDVAAESSEQPMVE
SITKLDGVPSADGVPESNQNAHQDEDDDDVTETNKVPEIIVLDGDDDDAAPP
LPKIEPPEPEIKVSEMELMGPPPTMTISTKTDFEAIKIPSPGILAMPPTLNNTN
NNNNNGSDNCKLTGLGQQQNFOSVKTEDVKEDDICIIVTSVSDTTPKW
STVRVRLPILSLPAEAGGVQVQELQISVANONCSAQLOQHPDNLNMOYYSIPM
DECKVDTSKLNGECIFSLGLDEKQMLAVKEKYSKNGLSGPERHETKDEE
QAKLKDIEDTMTDADDLAGKEKFLRSDVDPDTGCGVSGDVKPKLEKLE
ALQAYTHIANNSVSVQTYIPIIDIPLSMTPEHRLLEQVLEQVLPVGVVYVP
RQRYWMLDQKQWPKVAKRVSLALIEQLESLEDKVASASMLQNLQVYKLSKSID
LDSQEDVETEDVSIIPMIRERI IIDEANIRRYLKPILGSGTGAHLIAQNHQHT
TOTONSASAYLLQWQQQQQQQOQQQQQQQGGAGNSLNPSSFNERTWALAAA
AASGPNATGVANSVAGATCEGSGEPNSGNASPASNCDSRDEKVEQIPLGLVQ
WDAVSRHTTQALAMALVLESYVWAKSINKAYATKNSKSKSSAKQATPSKKQ
QCKNKKQQLTPNGESKSKSAINKPENQAPLSIKILKLAQNGQCLLTKTPIL
TASPSPSSHPNTSDSDFGKRTKSGSKRRRSANNTSSKYSNLQNCQFQTS
ENEDKLLCDGCKGYTYCFKPMNDIPDGWYCYBCYNATKRCIKVCGGHRPSP
VGKMTYCDLCPHADCYIPLLVPRGWYCHGICISRAPPKRSAGTSGSSKS
RRDRSGSGSKRSKTSKPMESHWQQQWPLAGGDSHHHHQCPPLSNASHDESM
NSLPAGSSSVSGACGLNGRESITSGQLNNASALCSPAESVVSATYDDQHHANS

```

FEATURES
source

```

VDGSRFHAHLIPPSNNGTAALLDVEDVPGGANVPGYVYVTVAAAGNFSAGLINQAPV
QPAMPFANVAMSRAVYTRTRKTPPTAPTPPPPPPTLLMQASPTATALLHVAACQ
SPPOOHAQIATMPSPPAIGATNQMSPPINIHAIQAEKIKQEKKGKATKKL
MKELAVCKTLLGEMHEDSWPFLPVNTKQFTYRKIIKTPMDLSITKKQLQDLSYK
TREDPCVDVRQIFDNCENFEDDSPVGAGHGWRKXFFESRWGELTDKHS"

Query Match      81.7%; Score 18.8; DB 3; Length 39574;
Best Local Similarity 90.9%; Pred. No. 6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTGGTT 22
      |||||
Db 13550 TAAATCATATGCGTTTGGTT 13571

RESULT 13
AC020320
LOCUS      AC020320              74881 bp      DNA      linear      HTC 03-JAN-2000
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
ACCESSION  AC020320
VERSION     AC020320.1  GI:6664577
KEYWORDS    HTG; HTGS PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 74881)
AUTHORS     Adams,M. and Venter,J.C.
TITLES      Direct Submission
JOURNAL     Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212756 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES     Location/Qualifiers
             source
              1..74881
              /organism="Drosophila melanogaster"
              /mol_type="genomic DNA"
              /db_xref="taxon:7227"

Query Match      81.7%; Score 18.8; DB 2; Length 74881;
Best Local Similarity 90.9%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTGGTT 22
      |||||
Db 37421 TAAATCATATGCGTTTGGTT 37442

RESULT 14
AC010677/c
LOCUS      AC010677              112659 bp      DNA      linear      PRI 03-JAN-2002
DEFINITION Homo sapiens chromosome 7 clone CTD-2304L4, complete sequence.
ACCESSION  AC010677
VERSION     AC010677.4  GI:11465112
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 112659)
AUTHORS     Waterston,R.H.
TITLES      The sequence of Homo sapiens clone
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 112659)
AUTHORS     Waterston,R.H.
TITLES      Direct Submission

```

JOURNAL Submitted (17-SEP-1999) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 112659)
AUTHORS Waterston,R.H.
TITLES Direct Submission
JOURNAL Submitted (30-NOV-2000) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 112659)
AUTHORS Waterston,R.
TITLES Direct Submission
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 112659)
AUTHORS Waterston,R.H.
TITLES Direct Submission
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT On Nov 30, 2000 this sequence version replaced gi:7630788.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submission@wustl.wustl.edu
----- Project Information -----
Center project name: H.MS2304L04

FEATURES
source Location/Qualifiers
1..112659
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/clone="CTD-2304L4"
ORIGIN
Query Match 81.7%; Score 18.8; DB 9; Length 112659;
Best Local Similarity 90.9%; Pred.No. 4.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TTAATCATATCGCTTTTGGTT 22
||||| ||||| ||||| ||||| |||||
DB 97079 TTAATCAATGCGCTTTTGGTT 97058

RESULT 15
BX005186/c
LOCUS zebrafish DNA sequence from clone DKEY-11N4 in linkage group 21, complete sequence.
DEFINITION BX005186 119595 bp. DNA linear VRT 17-SEP-2003
ACCESSION BX005186
VERSION BX005186.9 GI:34850144
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 119595)
AUTHORS Smith,M.
TITLES Direct Submission
JOURNAL Submitted (17-SEP-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Sep 18, 2003 this sequence version replaced gi:32398550.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EMBL, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C.elegans/wormpep/Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.
Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhairong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-11N4 is from a Zebrafish BAC library
VECTOR: pIndigoBAC-5.

FEATURES
source Location/Qualifiers
1..119595
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-11N4"
/clone_lib="DanioKey"
ORIGIN
Query Match 81.7%; Score 18.8; DB 5; Length 119595;
Best Local Similarity 90.9%; Pred.No. 4.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TTAATCATATCGCTTTTGGTT 22
||||| ||||| ||||| ||||| |||||
DB 34624 TTAATCGTATCGCTTTTGGTT 34603
Search completed: May 26, 2004, 17:22:21
Job time : 324.936 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 15:20:44 ; Search time 292.115 Seconds
(without alignments)
3115.905 Million cell updates/sec

Title: US-10-676-299-8

Perfect score: 21

Sequence: 1 taacacaaacgcataatgatt 21

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hlg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pi.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_un.*
- 28: em_vt.*
- 29: em_vt.*
- 30: em_hlg_hum.*
- 31: em_hlg_inv.*
- 32: em_hlg_other.*
- 33: em_hlg_mus.*
- 34: em_hlg_pln.*
- 35: em_hlg_rod.*
- 36: em_hlg_man.*
- 37: em_hlg_vrt.*
- 38: em_sy.*
- 39: em_hlg_hum.*
- 40: em_hlg_mus.*
- 41: em_hlg_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	21	100.0	727	1	ECRFARSR	X16045 E. coli R-f
C 2	21	100.0	120826	1	AP005147	AP005147 Salmonella
C 3	18.4	87.6	21479	3	CBRG03B09	AC084483 Caenorhab
C 4	18.4	87.6	34796	3	CEFS3C11	279756 Caenorhabdi
C 5	18.4	87.6	39574	3	AF314193	AF314193 Drosophila
C 6	18.4	87.6	74881	2	AC020320	AC020320 Drosophila
C 7	18.4	87.6	170869	3	AC011696	AC011696 Drosophila
C 8	18.4	87.6	171831	3	AC007473	AC007473 Drosophila
C 9	18.4	87.6	261690	3	AE003825	AE003825 Drosophila
C 10	18	85.7	96444	9	HSJ738116	AL049867 Human DNA
C 11	18	85.7	119082	2	AC008057	AC008057 Homo sapi
C 12	18	85.7	169494	9	AC007630	AC007630 Homo sapi
C 13	17.8	84.8	466	3	BOC506884	AJ506884 Butus oc
C 14	17.8	84.8	466	3	BOC506893	AJ506893 Butus oc
C 15	17.8	84.8	466	3	BOC506894	AJ506894 Butus oc
C 16	17.8	84.8	466	3	BOC506896	AJ506896 Butus oc
C 17	17.8	84.8	466	3	BOC506897	AJ506897 Butus oc
C 18	17.8	84.8	466	3	BOC506901	AJ506901 Butus oc
C 19	17.8	84.8	466	3	BOC506902	AJ506902 Butus oc
C 20	17.8	84.8	491	3	MEU550701	AJ550701 Mesobuthu
C 21	17.8	84.8	491	3	MEU550702	AJ550702 Mesobuthu
C 22	17.8	84.8	491	3	MEU550703	AJ550703 Mesobuthu
C 23	17.8	84.8	599	3	AY128129	AY128129 Halgerda
C 24	17.8	84.8	599	3	AY128138	AY128138 Halgerda
C 25	17.8	84.8	603	3	MTAJ3267	AJ23267 Jorunna t
C 26	17.8	84.8	658	3	AF370829	AF370829 Androcton
C 27	17.8	84.8	676	3	AY116607	AY116607 Styela cl
C 28	17.8	84.8	778	5	AY169861	AY169861 Morelia v
C 29	17.8	84.8	1263	3	S54796	S54796 cytochrome
C 30	17.8	84.8	10274	1	U32797	U32797 Haemophilus
C 31	17.8	84.8	14771	3	AB024528	AB024528 Halocynth
C 32	17.8	84.8	110000	2	EX005451_0	EX005451 Mus muscu
C 33	17.8	84.8	110000	6	AR274513_12	Continuation (13 o
C 34	17.8	84.8	110220	2	AP003688	AP003688 Oryza sat
C 35	17.8	84.8	127887	2	AC125477	AC125477 Medicago
C 36	17.8	84.8	131402	10	AL773583	AL773583 Mouse DNA
C 37	17.8	84.8	137635	2	EX629344	EX629344 Danio rer
C 38	17.8	84.8	162297	8	AP004317	AP004317 Oryza sat
C 39	17.8	84.8	165679	8	AC125475	AC125475 Medicago
C 40	17.8	84.8	177140	2	AP005612	AP005612 Oryza sat
C 41	17.8	84.8	233016	2	AC113719	AC113719 Rattus no
C 42	17.8	84.8	254644	2	AC136836	AC136836 Rattus no
C 43	17.8	84.8	256511	2	AC135040	AC135040 Rattus no
C 44	17.8	84.8	262721	2	AC106265	AC106265 Rattus no
C 45	17.8	84.8	270121	2	AC132060	AC132060 Rattus no

ALIGNMENTS

RESULT 1	ECRFARSR	727 bp	DNA	linear	BCT 07-SEP-1994
ECRFARSR/c	ECRFARSR	727 bp	DNA	linear	BCT 07-SEP-1994
LOCUS	E.coli R-factor R773 arsr gene.				
DEFINITION	X16045				
ACCESSION	X16045.1	GI:42716			
VERSION	arsenal resistance; arsr gene; Arsr protein; DNA-binding protein;				
KEYWORDS	regulatory protein; resistance gene.				
SOURCE	Escherichia coli				
ORGANISM	Escherichia coli				
	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
	Enterobacteriaceae; Escherichia.				
REFERENCE	1 (bases 1 to 727)				
AUTHORS	San Francisco M.J., Hope, C.L., Owolabi, J.B., Tisa, L.S. and Rosen, B.P.				

Pred. No. is the number of results predicted by chance to have a

	JOURNAL	MEDLINE	MOL. GEN. GENET.	213 (1), 30-35 (1988)
	PUBLISHED	89127142		
	REFERENCE	3065610		
	AUTHORS	Komano,T., Toyoshima,A., Morita,K. and Nishioka,T.		
	TITLE	Cloning and nucleotide sequence of the orf region of the IncII plasmid R64		
	JOURNAL	J. Bacteriol.	170 (9), 4385-4387 (1988)	
	PUBLISHED	88314948		
	REFERENCE	3045094		
	AUTHORS	Furuya,N., Nishioka,T. and Komano,T.		
	TITLE	Nucleotide sequence and functions of the orf operon in IncII plasmid R64		
	JOURNAL	J. Bacteriol.	173 (7), 2231-2237 (1991)	
	PUBLISHED	91177811		
	REFERENCE	1848841		
	AUTHORS	Furuya,N. and Komano,T.		
	TITLE	Determination of the nick site at orit of IncII plasmid R64: global similarity of orit structures of IncII and IncP plasmids		
	JOURNAL	J. Bacteriol.	173 (20), 6612-6617 (1991)	
	PUBLISHED	92011438		
	REFERENCE	1917882		
	AUTHORS	Kim,S.R., Funayama,N. and Komano,T.		
	TITLE	Nucleotide sequence and characterization of the traABCD region of IncII plasmid R64		
	JOURNAL	J. Bacteriol.	175 (16), 5035-5042 (1993)	
	PUBLISHED	93352408		
	REFERENCE	8349545		
	AUTHORS	Furuya,N. and Komano,T.		
	TITLE	Surface exclusion gene of IncII plasmid R64: nucleotide sequence and analysis of deletion mutants		
	JOURNAL	Plasmid	32 (1), 80-84 (1994)	
	PUBLISHED	95083745		
	REFERENCE	7991676		
	AUTHORS	Furuya,N. and Komano,T.		
	TITLE	Nucleotide sequence and characterization of the trbABC region of the IncII plasmid R64: existence of the pnd gene for plasmid maintenance within the transfer region		
	JOURNAL	J. Bacteriol.	178 (6), 1491-1497 (1996)	
	PUBLISHED	96198148		
	REFERENCE	8626273		
	AUTHORS	Kim,S.R. and Komano,T.		
	TITLE	The plasmid R64 thin pilus identified as a type IV pilus		
	JOURNAL	J. Bacteriol.	179 (11), 3594-3603 (1997)	
	PUBLISHED	97315231		
	REFERENCE	9171405		
	AUTHORS	Narshara,K., Rahman,E., Furuya,N. and Komano,T.		
	TITLE	Requirement of a limited segment of the sog gene for plasmid R64 conjugation		
	JOURNAL	Plasmid	38 (1), 1-11 (1997)	
	PUBLISHED	97428559		
	REFERENCE	9281491		
	AUTHORS	Furuya,N. and Komano,T.		
	TITLE	Mutational analysis of the R64 orit region: requirement for precise location of the Nika-binding sequence		
	JOURNAL	J. Bacteriol.	179 (23), 7291-7297 (1997)	
	PUBLISHED	98053841		
	REFERENCE	9393692		
	AUTHORS	Yoshida,T., Furuya,N., Ishikura,M., Isobe,T., Haino-Fukushima,K., Ogawa,T. and Komano,T.		
	TITLE	Purification and characterization of thin pilus of IncII plasmids Colib-P9 and R64: formation of PilV-specific cell aggregates by type IV pilus		
	JOURNAL	J. Bacteriol.	180 (11), 2842-2848 (1998)	
	PUBLISHED			
	REFERENCE			
	AUTHORS			
	TITLE			
	JOURNAL			
	PUBLISHED			
	REFERENCE			
	AUTHORS			
	TITLE			
	JOURNAL			
	PUBLISHED			
	REFERENCE			
	AUTHORS			
	TITLE			
	JOURNAL			
	PUBLISHED			
	REFERENCE			
	AUTHORS			
	TITLE			
	JOURNAL			
	PUBLISHED			
	REFERENCE			
	AUTHORS			
	TITLE			
	JOURNAL			
	PUBLISHED			
	REFERENCE			
	AUTHORS			
	TITLE			
	JOURNAL			
	PUBLISHED			
	REFERENCE			
	AUTHORS			
	TITLE			
	JOURNAL			
	PUBLISHED			
	REFERENCE			
	AUTHORS			
	TITLE			
	JOURNAL			
	PUBLISHED			
	REFERENCE			
	AUTHORS			
	TITLE			
	JOURNAL			
	PUBLISHED			
	REFERENCE			
	AUTHORS			
	TITLE			
	JOURNAL			
	PUBLISHED			
	REFERENCE			
	AUTHORS			
	TITLE			
	JOURNAL			
	PUBLISHED			
	REFERENCE			
	AUTHORS			
	TITLE			
	JOURNAL			
	PUBLISHED			
	REFERENCE			
	AUTHORS			
	TITLE			
	JOURNAL			
	PUBLISHED			
	REFERENCE			

```

MEDLINE      98268996
PUBMED
REFERENCE    9603870
AUTHORS      Komano,T., Yoshida,T., Narahara,K. and Furuya,N.
TITLE        The transfer region of IncII plasmid R64: similarities between R64
              tra and legionella icm/dot genes
JOURNAL      Mol. Microbiol. 35 (6), 1348-1359 (2000)
MEDLINE      20223621
PUBMED
REFERENCE    10760136
AUTHORS      Sampei,G., Komano,T., Sasaki,T., Tachibana,K., Furuya,N., Saito,Y.,
              Suzuki,T. and Mizobuchi,K.
TITLE        Organisation and diversification of plasmid genomes: complete
              nucleotide sequence of R64 genome
JOURNAL      Direct Submission
REFERENCE    15 (bases 1 to 120826)
AUTHORS      Komano,T., Sasaki,T., Tachibana,K., Furuya,N., Saito,Y., Suzuki,T.,
              Sampei,G. and Mizobuchi,K.
TITLE        Direct Submission
JOURNAL      Submitted (07-MAY-2002) Gen-ichi Sampei, The University of
              Electro-Communications, Applied Physics and Chemistry; Chofugaoka
              1-5-1, Chofu, Tokyo 182-8585, Japan [E-mail: sampei@pc.uec.ac.jp,
              Tel:81-424-43-5481, Fax:81-424-43-5501]
FEATURES     Location/Qualifiers
              source
                1..120826
                /organism="Salmonella typhimurium"
                /mol_type="genomic DNA"
                /db_xref="taxon:602"
                /lab_host="Escherichia coli strain K-12"
                /plasmid="R64"
                /note="R64 strain drd-11"
              complement(join(20047..20418,3609..4147))
              /note="100 pct identical to sp:Y122_ECOLI[hypothetical
              protein of insertion sequence IS2]
              location complement(4143..4147) and location
              complement(20047..20051) are duplicated"
              378..467
              /gene="repY"
              378..467
              /gene="repY"
              /note="100 pct identical to pIR:A35445[RepY of plasmid
              Colib-p9]
              possible regulatory reading frame"
              /codon_start=1
              /transl_table=11
              /protein_id="BAB91567.1"
              /db_xref="GI:20521503"
              /translation="MKPYQRFNPVQCINFRNRSAISDSLMQV"
              455..1486
              /gene="repZ"
              455..1486
              /gene="repZ"
              /note="100 pct identical to sp:REPZ_ECOLI,sp:P18023[RepZ
              of plasmid Colib-p9]"
              /codon_start=1
              /transl_table=11
              /protein_id="BAB91568.1"
              /db_xref="GI:20521504"
              /translation="NAGLNTFNVAVHNSQLAPEQIRFWEDYEGRAFTFLVPEPERK
              RYKRGERTKPKCNPSWYRPERYKALKGQGHAYNELVKKQPVTCQSRLMRMSR
              HFYQYKRTFVRKYNAPREKGRLLDAIWPVLSFSDAGTHTVGMSTFLAEISPKD
              SEGHVTPLEVTVSRLALQVRFGLGVSEETMWDREHQRQLRPRIWITPAGNQM
              LGVDMVKLEHQQKRLSEIRIQQLIREGVLEDSDISVHAARKKWTYLSQDALKR
              REKAASKKANLKKLPVDQQTYYEAYELRKLPDPBAYFCSDDLRLRLAIRELQLE
              LTLAAPPPH"
              2761..3114
              /gene="arsR"
              2761..3114
              /gene="arsR"
              /note="100 pct identical to sp:ARR1_ECOLI,sp:P15905[arsR
              of plasmid R773]"
              /codon_start=1
              /transl_table=11
protein_id="BAB91569.1"
/db_xref="GI:20521505"
/translation="MLQITPLQFKNLSDETRLGIVLLIREMGLCVCDLQWALDQSQ
PKISRLAMRESGILLDRQKGVHYRLSPHPSWAAQIIEQAWLSQDDVQVITARK
LASVNCSSSKAVCI"
3162..3524
/gene="arsD"
3162..3524
/gene="arsD"
/note="99 pct identical to sp:ARD1_ECOLI,sp:P46003[arsD of
plasmid R773]"
/codon_start=1
/transl_table=11
/protein_id="BAB91570.1"
/db_xref="GI:20521506"
/translation="MKTLVFDPMCCSTGVCGTVDQALVDFSDVQWLKQCGVQIE
RFNLAQPMFQVQNEKVKAFTEASGAGLPIILLDGETVMAGRYPKRABLARWFGIPL
DKVGLAPSGCGGNTSC"
join(3542..3597,20833..22533)
/gene="arsA"
join(3542..3597,20833..22533)
/gene="arsA"
/note="99 pct identical to pIR:A25937[arsA of plasmid
R773, arsenical pump-driving ATPase]
location 3593..3597 and location 20833..20837 are
duplicated"
join(3598..4147,20047..20832)
/insertion_seq="IS2"
complement(join(3609..4147,20047..20418))
/gene="yadA"
join(4148..7818,18208..20046)
/note="different from Tn5393[gb:M96392] in IS1133 insert
position and terminal direct repeat"
Query Match 100.08; Score 21; DB 1; Length 120826;
Best Local Similarity 100.08; Pred. No. 23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAACCAAAAACGCATATCATTT 21
|||||
DB 2700 TAACCAAAAACGCATATCATTT 2690

RESULT 3
CBRG03E09
LOCUS       CBRG03E09          21479 bp    DNA     linear     INV 04-NOV-2000
DEFINITION Caenorhabditis briggsae cosmid G03E09, complete sequence.
ACCESSION   AC084483
VERSION     AC084483.1   GI:21094933
SOURCE      HTG.
            Caenorhabditis briggsae
            Caenorhabditis briggsae
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
ORGANISM    1 (bases 1 to 21479)
            Washington University Genome Sequencing Center.
            The C. briggsae Genome Sequencing Project
            Unpublished
REFERENCE    2 (bases 1 to 21479)
            Waterston,R.
            Direct Submission
            Submitted (04-NOV-2000) Department of Genetics, Washington
            University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            Submitted by:
            Genome Sequencing Center
            Department of Genetics, Washington University,
            St. Louis, MO 63110, USA
            e-mail: jspie@wustl.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between

```

```

FEATURES
  source
    neighboring submissions.
    Location/Qualifiers
      1. .21479
      /organism="Caenorhabditis briggsae"
      /mol_type="genomic DNA"
      /strain="Gujarat G16"
      /db_xref="taxon:6238"
      /clone="G03E09"

ORIGIN
  Query Match      87.6%; Score 18.4; DB 3; Length 21479;
  Best Local Similarity 95.0%; Pred. No. 4.6e+02;
  Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

  QY 1 TAACCAAAAACGCATATGAT 20
      |||||
  Db 12998 TAACCAAAAACGCATATGTT 13017

RESULT 4
CEP53C11
LOCUS CEP53C11 34796 bp DNA linear INV 03-DEC-2003
DEFINITION Caenorhabditis elegans cosmid F53C11, complete sequence.
ACCESSION Z79756
VERSION Z79756.1 GI:1523939
KEYWORDS HTG; 2,4-dienoyl-CoA reductase; Potassium channel protein like;
          Yeast hypotheical protein YPL247C like.
SOURCE Caenorhabditis elegans
ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
          Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.
REFERENCE 1
AUTHORS none.
TITLE Genome sequence of the nematode C. elegans: a platform for
        investigating biology. The C. elegans Sequencing Consortium
        Science 282 (5396), 2012-2018 (1998)
JOURNAL 99069613
MEDLINE 9851916
PUBMED 9851916
REMARK The C. elegans Sequencing Consortium.
REFERENCE 2 (bases 1 to 34796)
AUTHORS Baynes, C.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-1996) Nematode Sequencing Project, Sanger
        Institute, Hinxton, Cambridge CB10 1SA, England and Department of
        Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
        jess@sanger.ac.uk or rw@nemastode.wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using
        predictions from Genefinder (P. Green, U. Washington), and other
        available information.
        Current sequence finishing criteria for the C. elegans genome
        sequencing consortium are that all bases are either sequenced
        unambiguously on both strands, or on a single strand with both a
        dye primer and dye terminator reaction, from distinct subclones.
        Exceptions are indicated by an explicit note.
        For a graphical representation of this sequence and its analysis
        see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
        name=F53C11
        IMPORTANT: This sequence is NOT necessarily the entire insert of
        the specified clone. It may be shorter because we only sequence
        overlapping sections once, or longer because we arrange for a small
        overlap between neighbouring submissions.
        IMPORTANT: This sequence is not the entire insert of clone F53C11.
        It may be shorter because we only sequence overlapping sections
        once, or longer because we arrange for a small overlap between
        neighbouring submissions.
        The true left end of clone F53C11 is at 1 in this sequence. The
        true right end of clone F53C11 is at 10263 in
        sequence CEP55B12.
        The true left end of clone F55B12 is at 34695 in this sequence. The
        start of this sequence (1..100) overlaps with the end of sequence
        CER2002.
        The end of this sequence (34695..34796) overlaps with the start of
        sequence CEP55B12.

FEATURES
  source
    Location/Qualifiers
      1. .34796
      /organism="Caenorhabditis elegans"
      /mol_type="genomic DNA"
      /strain="Bristol N2"
      /db_xref="taxon:6233"
      /clone="F53C11"

ORIGIN
  Query Match      87.6%; Score 18.4; DB 3; Length 34796;
  Best Local Similarity 95.0%; Pred. No. 4.2e+02;
  Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

  QY 2 AACCAAAAACGCATATGATT 21
      |||||
  Db 221 AACCAAAAACGCATATATT 240

RESULT 5
AF314193/c
LOCUS AF314193 39574 bp DNA linear INV 11-APR-2003
DEFINITION Drosophila melanogaster Toutatis (tou) gene, complete cds.
ACCESSION AF314193
VERSION AF314193.1 GI:12642597
KEYWORDS Drosophila melanogaster (fruit fly)
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
          Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 39574)
AUTHORS Fauvarque, M.O., Laurenti, P., Bolvin, A., Bloyer, S., Griffin-Shea, R.,
          Bourbon, H.M. and Dura, J.M.
          Dominant modifiers of the polyhomeotic extra-sex-combs phenotype
          induced by marked P element insertion mutagenesis in Drosophila
          Genet. Res. 78 (2), 137-148 (2001)
JOURNAL 21589546
MEDLINE 11732091
PUBMED 11732091
REMARK The novel trx-G gene toutatis encodes a Drosophila member of the
          WAL family of chromatin-associated proteins
          Unpublished
          3 (bases 1 to 39574)
          Bourbon, H.M.
          Direct Submission
          TITLE
          JOURNAL
          REFERENCE
          AUTHORS
          TITLE
          JOURNAL
          COMMENT Submitted (18-OCT-2000) Centre de Biologie du Developpement,
          CNRS/UPS, 118 Route de Narbonne, Toulouse, HG 31062, France

FEATURES
  source
    Location/Qualifiers
      1. .39574
      /organism="Drosophila melanogaster"
      /mol_type="genomic DNA"
      /db_xref="taxon:7227"
      /chromosome="2"
      /map="4BA3-6"
      /gap="1651..38602"
      /gene="tou"
      /join(1651..1903,25307..25489,25552..25797,26308..26795,
      26857..27479,27542..28784,28847..29119,29355..29711,30457..34467,
      30457..34467,35337..35854,36084..37248,37313..37434,
      37500..38602)
      /gene="tou"
      /product="Toutatis"
      /join(25363..25489,25552..25797,26308..26795,26857..27479,
      27542..28784,28847..29119,29355..29711,30457..34467,
      35337..35854,36084..37248,37313..37434,37500..37656)
      /gene="tou"
      /note="putative subunit of a chromatin-remodeling complex"
      /codon_start=1
      /product="Toutatis"
      /protein_id="AAK00302.1"
      /db_xref="GI:12642598"
      /translation="MNKAGDGS DGNKNSKNSNAGGGGAGGPHDPTGLDASLFAY"

```

<hr/>					
ORIGIN					
<hr/>					
Query Match	87.6%	Score 18.4;	DB 3;	Length 39574;	
Best Local Similarity	95.0%;	Pred. No.	4.1e+02;		
Matches 19;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;	
<hr/>					
OY	2	AACCAAAACGCATATGATT	21		
Dd	13571	AACCCAAAACGCATATGATT	13552		
<hr/>					
RESULT 6					
ACO20320/c					
LOCUS			74881 bp	DNA	linear HTG 03-JAN-2000
DEFINITION					*** SEQUENCING IN PROGRESS ***.
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					(fruit fly)
ORGANISM					
					Drosophila melanogaster
					Drosophila melanogaster
					HTG; HTGS_PHASE2.
					GI:6664577
					ACO20320.1
					Drosophila melanogaster,

JOURNAL REFERENCE AUTHORS	Unpublished 2 (bases 1 to 170869)	TITLE
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoakins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomocan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequelira, A., Sethi, H., Snit, E., Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.	2 (bases 1 to 170869)	Direct Submission
		Submitted (11-OCT-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
		On Mar 23, 2001 this sequence version replaced gi:16119492.

Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdgp@fruitfly.berkeley.edu.

Location/Qualifiers
 1. 170869
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /strain="y; cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="2R"
 /map="48A-48C"
 /clone="BACR35F01 (D1156)"
 /clone_lib="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6)"

ORIGIN
 Query Match 87.6%; Score 18.4; DB 3; Length 170869;
 Best Local Similarity 95.0%; Pred. No. 3.2e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AACCAAAAGCGCATATGATT 21
 |||||
 DB 20114 AACCAAAAGCGCATATGATT 20133

RESULT 8
 AC007473

LOCUS
 Drosophila melanogaster, chromosome 2R, region 48A-48B, BAC clone BACR38D12, complete sequence.

ACCESSION
 AC007473

VERSION
 AC007473.10 GI:13162476

KEYWORDS
 HTG

SOURCE
 Drosophila melanogaster (fruit fly)

ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
 1 (bases 1 to 171831)
 Cealiker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C., Rogers, Y., An, H., Balgwin, D., Banzon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferreira, S., Frise, E., Galle, R.F., Gaig, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Thegwan, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

TITLE
 Sequencing of Drosophila chromosome 2R, region 48A-48B

Unpublished

JOURNAL
 2 (bases 1 to 171831)

REFERENCE
 Cealiker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhorff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.B., Galle, R., George, R.A., Harris, N.I., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Serueira, A., Sethi, H., Snir, E., Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.B. and Rubin, G.M.

TITLE
 Direct Submission

JOURNAL
 Submitted (05-MAY-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

COMMENT
 On Feb 28, 2001 this sequence version replaced gi:5670618.

Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdgp@fruitfly.berkeley.edu.

Location/Qualifiers
 1. 171831
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /strain="y; cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="2R"
 /map="48A-48B"
 /clone="BACR38D12 (D590)"
 /clone_lib="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6)"

ORIGIN
 Query Match 87.6%; Score 18.4; DB 3; Length 171831;
 Best Local Similarity 95.0%; Pred. No. 3.2e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AACCAAAAGCGCATATGATT 21
 |||||
 DB 60433 AACCAAAAGCGCATATGATT 60452

RESULT 9
 AC003825

LOCUS
 Drosophila melanogaster chromosome 2R, section 25 of 74 of the complete sequence.

ACCESSION
 AC003825

VERSION
 AC003825.3 GI:21627432

KEYWORDS
 SOURCE

ORGANISM
 Drosophila melanogaster (fruit fly)

Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
 1 (bases 1 to 261690)
 Adams, M.D., Celisner, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pranckoch, C., Balgwin, D., Balow, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablo, B., Delcher, A., Deng, Z., Dwyer, I., Dietz, S.N., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferreira, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garb, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwan, C., Jalali, M., Kalush, P., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C.,


```
SVGQCSBAQAALSAQARQRLISQIYRPSASSTATTVLPSGEGPPSPEDLLQ
LPPSTGTOSEIRKSYAEELMQQMLMAARVNALTAAGKQQLQMAAAVA
TPVSDALAQITAGLPGGAVHPHQDQLLQDQVHHHMHQHLNNHHERA
LKFSDNLALQDGLPKIGSLGNIQSGSVSGSTGSKNSGNRNPRLKPKK
SKPILNLAQNAARSLSFSSLANICSNNSNSTATSTSTTSAGPDLVKSP
PAAGATGSKSGEDSGTPIVPAWVCTYSDRPSGSGSTAKAAGADVPQGR
WNGWRGSGGGRGAGGAGGAAAGLORNAVGTARVORESLSDGATPAERGTG
TERGADQLVPQTQAEKVERHQESAGNAADGAGIVQPLDDTADPRGGAGAGG
"
CDS
Join(10671..11968,40764..40924,41344..41585)
/genes="inv"
/locus_tag="CG17835"
/note="inv gene product from transcript CG17835-RC"
/codon_start=1
/product="CG17835-PC"
/protein_id="AA068707.2"
/db_xref="GI:23240330"
/db_xref="FlyBase:FBgn001269"
/translation="NSTLASTRPPLKUTIPSLSEADHAQERRAGGQGVGRKHPD
Query Match 87.6%; Score 18.4; DB 3; Length 261690;
Best Local Similarity 95.0%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 AACCAAAACGCATATGATT 21
Db 138606 AACCAAAACGCATATGATT 138625
RESULT 10
HSJ738116 96444 bp DNA linear PRI 19-MAR-2001
LOCUS Human DNA sequence from clone RP4-738116 on chromosome
20p11.22-12.2. Contains STSs and GSSs, complete sequence.
ACCESSION AL049867
VERSION AL049867.2 GI:5002693
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLES Bates,K.
JOURNAL Direct Submission
Submitted (14-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: Clonerequest@sanger.ac.uk
On Jun 7, 1999 this sequence version replaced gi:4837713.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30). An attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone, and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Swi,
SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
Chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RP4-738116 is from the library RP4-4 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
```

```
VECTOR: PCYPAC2
IMPORTANT: This sequence is not the entire insert of clone
RP4-738116. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP4-738116 is at 96444 in this
sequence. The true right end of clone RP5-106901 is at 101 in this
sequence.
FEATURES
Location/Qualifiers
1..96444
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="p11.22-12.2"
/clone="RP4-738116"
/clone_lib="RPCI-4"
1590..1722
/repeat_region
/note="FLAM A repeat: matches 1..132 of consensus"
2014..2082
/repeat_region
/note="THR1B repeat: matches 3..71 of consensus"
3083..3383
/misc_feature
/note="match: GSS: Em:AQ245299"
3384..3689
/repeat_region
/note="AluSq repeat: matches 1..303 of consensus"
4413..4478
/repeat_region
/note="L1ME repeat: matches 5387..5456 of consensus"
4512..4873
/repeat_region
/note="THR1B repeat: matches 1..364 of consensus"
5024..5278
/repeat_region
/note="L1MC5 repeat: matches 7084..7338 of consensus"
5279..5715
/repeat_region
/note="MER83 repeat: matches 1..448 of consensus"
5716..5904
/repeat_region
/note="L1MC5 repeat: matches 7338..7526 of consensus"
7635..7821
/repeat_region
/note="WIR repeat: matches 16..209 of consensus"
8314..8542
/repeat_region
/note="WIR repeat: matches 21..259 of consensus"
9835..10001
/repeat_region
/note="MER5B repeat: matches 1..175 of consensus"
10017..10149
/repeat_region
/note="MERSA repeat: matches 10..174 of consensus"
11149..11276
/repeat_region
/note="FLAM C repeat: matches 4..133 of consensus"
11590..11816
/repeat_region
/note="L2 repeat: matches 1221..1435 of consensus"
11821..11901
/repeat_region
/note="WIR repeat: matches 12..97 of consensus"
11908..11941
/repeat_region
/note="17 copies 2 mer ac 91% conserved"
12082..12140
/repeat_region
/note="L1ME repeat: matches 5733..5791 of consensus"
12172..12213
/repeat_region
/note="L1P5 repeat: matches 5617..5657 of consensus"
12244..12334
/repeat_region
/note="L1ME repeat: matches 5520..5610 of consensus"
12434..12606
/repeat_region
/note="WIR repeat: matches 53..262 of consensus"
13110..13258
/repeat_region
/note="L2 repeat: matches 2061..2219 of consensus"
13617..13720
/repeat_region
/note="52 copies 2 mer tt 60% conserved"
14025..14534
/repeat_region
/note="LTR10D repeat: matches 1..513 of consensus"
14648..14728
/repeat_region
/note="WIR repeat: matches 48..136 of consensus"
15204..15240
/misc_feature
/note="match: GSS: Em:AQ429911
match: STS: Em:A0025273"
15204..15238
/misc_feature
/note="match: STS: Em:A0025603 Em:A0047422"
complement(15204..15238)
/note="match: STS: Em:A0024963 Em:A0026718"
```

```

misc_feature      complement(15204..15237)
/misc_feature     /note="match: STS: Em:AU047995"
/misc_feature     complement(15204..15236)
/misc_feature     /note="match: STS: Em:G44780"
/misc_feature     15204..15235
/misc_feature     /note="match: STS: Em:AU026333"
/misc_feature     15204..15233
/misc_feature     /note="match: STS: Em:AU027699"
/misc_feature     15472..15643
/misc_feature     /note="LIMC4 repeat: matches 5730..5898 of consensus"
/misc_feature     15893..16113
/misc_feature     /note="MIR repeat: matches 17..260 of consensus"
/misc_feature     16353..16716
/misc_feature     /note="MSTB repeat: matches 1..424 of consensus"
/misc_feature     16997..17025
/misc_feature     /note="LIMC4 repeat: matches 7855..7977 of consensus"
/misc_feature     17088..17841
/misc_feature     /note="LIMC4 repeat: matches 7098..7838 of consensus"
/misc_feature     19141..19415
/misc_feature     /note="L2 repeat: matches 2405..2705 of consensus"
/misc_feature     19519..19746
/misc_feature     /note="MIR repeat: matches 2..236 of consensus"
/misc_feature     20814..20855
/misc_feature     /note="21 copies 2 mer aa 76% conserved"
/misc_feature     20945..21492
/misc_feature     /note="L1PA16 repeat: matches 5606..6157 of consensus"
/misc_feature     21763..22095
/misc_feature     /note="L1MC5 repeat: matches 7586..7915 of consensus"
/misc_feature     23220..22411
/misc_feature     /note="23 copies 4 mer tatg 63% conserved"
/misc_feature     23372..22411
/misc_feature     /note="20 copies 2 mer tg 80% conserved"
/misc_feature     22526..22609
/misc_feature     /note="42 copies 2 mer ta 69% conserved"
/misc_feature     22773..23076
/misc_feature     /note="AluY repeat: matches 1..303 of consensus"
/misc_feature     23673..23832
/misc_feature     /note="L1M4 repeat: matches -11..147 of consensus"
/misc_feature     23860..23991
/misc_feature     /note="AluJ repeat: matches 4..135 of consensus"
/misc_feature     24147..24565
/misc_feature     /note="L1M4 repeat: matches 346..771 of consensus"
/misc_feature     24574..25747
/misc_feature     /note="L1M4 repeat: matches 1336..2112 of consensus"
/misc_feature     25808..26801
/misc_feature     /note="L1M4 repeat: matches 2177..3186 of consensus"
/misc_feature     complement(26802..27091)
/misc_feature     /note="match: STS: Em:L16406"
/misc_feature     26831..26965
/misc_feature     /note="L1M4 repeat: matches 3292..3447 of consensus"
/misc_feature     27031..27088
/misc_feature     /note="29 copies 2 mer ct 75% conserved"
/misc_feature     27092..27510
/misc_feature     /note="MLT2CB repeat: matches 3..453 of consensus"
/misc_feature     27512..28992
/misc_feature     /note="L1M4 repeat: matches 3177..4628 of consensus"
/misc_feature     29001..29496
/misc_feature     /note="L1MB5 repeat: matches 5453..5979 of consensus"
/misc_feature     29497..29775
/misc_feature     /note="AluX repeat: matches 23..307 of consensus"
/misc_feature     29776..29980
/misc_feature     /note="L1MB5 repeat: matches 5979..6176 of consensus"
/misc_feature     30375..30482
/misc_feature     /note="MLT1J repeat: matches 93..199 of consensus"
/misc_feature     31336..31638
/misc_feature     /note="AluSc repeat: matches 1..302 of consensus"
/misc_feature     complement(31816..32346)
/misc_feature     /note="match: GSS: Em:AQ344412"
/misc_feature     36485..36879
/misc_feature     /note="MBR39 repeat: matches 13..400 of consensus"
/misc_feature     36902..37259
/misc_feature     /note="TMB1C repeat: matches 1..369 of consensus"
/misc_feature     37403..38028

repeat_region    /note="L2 repeat: matches 2101..2745 of consensus"
repeat_region    39004..39135
repeat_region    /note="FLAM C repeat: matches 1..127 of consensus"
repeat_region    39005..39175
repeat_region    /note="BC200 repeat: matches 2..160 of consensus"
repeat_region    39294..39381
repeat_region    /note="MER5A repeat: matches 53..126 of consensus"
repeat_region    39466..39824
repeat_region    /note="MER39b repeat: matches 13..384 of consensus"
repeat_region    39827..39885
repeat_region    /note="MER5A repeat: matches 133..189 of consensus"
repeat_region    42662..42844

Query Match      85.7%; Score 18; DB 9; Length 96444;
Best Local Similarity 100.0%; Pred. No. 5.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AACCAAAAACGCATATGA 19
Db 23366 AACCAAAAACGCATATGA 23383

RESULT 11
AC008057 119082 bp DNA linear HTG 11-OCT-2000
LOCUS Homo sapiens chromosome 20 clone P1-10263, WORKING DRAFT SEQUENCE,
DEFINITION 3 unordered pieces.
ACCESSION AC008057
VERSION AC008057.4 GI:10765024
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 119082)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL McCombie, W.R.
JOURNAL Human Genomic Sequence, Chromosome 20
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 119082)
AUTHORS McCombie, W.R.
JOURNAL Direct Submission
JOURNAL Submitted (17-JUL-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
COMMENT On Oct 11, 2000 this sequence version replaced gi:9954568.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 68531: contig of 68531 bp in length
* 68532 68631: gap of unknown length
* 68632 102563: contig of 33932 bp in length
* 102564 102663: gap of unknown length
* 102664 119082: contig of 16419 bp in length.
FEATURES
source
1..119082
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/clone="P1-10263"

ORIGIN
Query Match      85.7%; Score 18; DB 2; Length 119082;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AACCAAAAACGCATATGA 19

```

Db 85514 AACCAAAACGCATATGA 85497

RESULT 12
AC007630
LOCUS Homo sapiens chromosome 20 clone P1-7739, complete sequence.
DEFINITION AC007630.4 GI:11245571
ACCESSION HTG.
VERSION Homo sapiens (human)
SOURCE
ORGANISM

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 169494)
AUTHORS McCombie, W.R.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
REFERENCE 2 (bases 1 to 169494)
AUTHORS de la Bastide, M., Gnoj, L., Dedhia, N.N., Matero, A., Ning Huang, E.,
O'Shaughnessy, A., Preston, R., Rodriguez, M., Schutz, K., Shah, R.,
Shekher, M., Siegel, L., Swaby, I., Vill, D. and McCombie, W.R.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
REFERENCE 3 (bases 1 to 169494)
AUTHORS McCombie, W.R.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-2000) Cold Spring Harbor Laboratory
Tn10 Transposon removed.
REMARK 4 (bases 1 to 169494)
AUTHORS McCombie, W.R.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-2000) Cold Spring Harbor Laboratory
Tn10 Transposon removed.
REMARK On Nov 21, 2000 this sequence version replaced gi:5332384.
COMMENT Clone P1-7739 contains a 3 kb overlap with clone 48F5. The overlap
ends at base 2940. From position 154930-155350 the subclones were
single stranded and had single chemistry.
Location/Qualifiers
1. .169494
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/clone="P1-7739"

misc_feature
134563
/note="Site of Tn10 transposon insertion. Tn10
Transposon, totalling 1328bp, was removed from final
sequence."
/evidence=not_experimental

ORIGIN

Query Match 85.7%; Score 18; DB 9; Length 169494;
Best Local Similarity 100.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AACCAAAACGCATATGA 19
|||||
DB 41629 AACCAAAACGCATATGA 41646

RESULT 13
BOC506884/c
LOCUS Buthus occitanus mardochei partial mitochondrial COI gene for
DEFINITION BOC506884.1 GI:28865338
ACCESSION cytochrome oxidase subunit I, haplotype AG4a.
VERSION AJ506884
KEYWORDS

Query Match 84.8%; Score 17.8; DB 3; Length 466;
Best Local Similarity 90.5%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAACCAAAACGCATATGATT 21
|||||
DB 164 TAACCAAAACGCATATGATT 144

RESULT 14
BOC506893/c
LOCUS Buthus occitanus mardochei partial mitochondrial COI gene for
DEFINITION BOC506893.1 GI:28865356
ACCESSION cytochrome oxidase subunit I, haplotype HA1b.
VERSION AJ506893
KEYWORDS COI gene; cytochrome oxidase subunit I.
SOURCE mitochondrion Buthus occitanus mardochei (Moroccan scorpion)
ORGANISM Buthus occitanus mardochei
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
Buthoidea; Buthidae; Buthus.

REFERENCE
AUTHORS Gantenbein, B. and Largiadet, C.R.
TITLE The phylogeographic importance of the Strait of Gibraltar as a gene
flow barrier in terrestrial arthropods: a case study with the
scorpion Buthus occitanus as model organism
Mol. Phylogenet. Evol. 28 (1), 119-130 (2003)
JOURNAL 22687009
MEDLINE 12801475
PUBMED
AUTHORS Gantenbein, B.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-2002) Gantenbein B., ICAPB, University of
Edinburgh, West Mains Road, Edinburgh EH9 3JT, UNITED KINGDOM
Location/Qualifiers
1. .466
/organism="Buthus occitanus mardochei"
/organelle="mitochondrion"
/mol_type="genomic DNA"
/sub_species="mardochei"
/specimen_voucher="A. Scholl, 8-II-1999 (NHMBE, CH)"
/db_xref="taxon:6869"
/haplotype="AG4a"
/country="Morocco:Imousser valley, Agadir region"
1. .466
/gene="COI"
/length=466
/codon_start=3
/transl_table=5
/product="cytochrome oxidase subunit I"
/protein_id="CD45304.1"
/db_xref="GI:28865339"
TRANSLATION="DQVYVYVTAHAFVMIFFVMVPMIGFGNVLVPLMVGADPMF
PRNMMSFLLPAPFLLSALLESAGTGTVPPLSSLAHMGSSVDLTIFSLHL
AGVSSILGAINFIITINMSMGMTDKVPLFVMSVFVAVLLLSLPVLAGA"

ORIGIN

Query Match 84.8%; Score 17.8; DB 3; Length 466;
Best Local Similarity 90.5%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAACCAAAACGCATATGATT 21
|||||
DB 164 TAACCAAAACGCATATGATT 144

RESULT 14
BOC506893/c
LOCUS Buthus occitanus mardochei partial mitochondrial COI gene for
DEFINITION BOC506893.1 GI:28865356
ACCESSION cytochrome oxidase subunit I, haplotype HA1b.
VERSION AJ506893
KEYWORDS COI gene; cytochrome oxidase subunit I.
SOURCE mitochondrion Buthus occitanus mardochei (Moroccan scorpion)
ORGANISM Buthus occitanus mardochei
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
Buthoidea; Buthidae; Buthus.

REFERENCE
AUTHORS Gantenbein, B. and Largiadet, C.R.
TITLE The phylogeographic importance of the Strait of Gibraltar as a gene
flow barrier in terrestrial arthropods: a case study with the
scorpion Buthus occitanus as model organism
Mol. Phylogenet. Evol. 28 (1), 119-130 (2003)
JOURNAL 22687009
MEDLINE 12801475
PUBMED
AUTHORS Gantenbein, B.
TITLE Direct Submission

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 15:20:44 ; Search time 347.756 Seconds
(without alignments)
3115.905 Million cell updates/sec

Title: US-10-676-299-9
Perfect score: 25
Sequence: 1 ttaagtcataatgttttgactta 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBankl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sta:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sta:*
28: em_un:*
29: em_vi:*
30: em_hgt_hum:*
31: em_hgt_inv:*
32: em_hgt_other:*
33: em_hgt_mus:*
34: em_hgt_pln:*
35: em_hgt_rod:*
36: em_hgt_mam:*
37: em_hgt_vrt:*
38: em_sv:*
39: em_hgt_hum:*
40: em_hgt_mus:*
41: em_hgt_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	25	100.0	3492	1	ECARSRBC	X80057 E.coli gene
2	25	100.0	10240	1	AE000426	AE000426 Escherich
3	25	100.0	11524	1	AE015361	AE015361 Shigella
4	25	100.0	179941	2	AC145934	AC145934 Gallus ga
5	25	100.0	225419	1	EC00076	U00039 E. coli chr
6	25	100.0	242495	2	AC146183	AC146183 Pan trogl
7	25	100.0	289816	1	AE016992	AE016992 Shigella
8	21.8	87.2	11071	1	AE005575	AE005575 Escherich
9	21.8	87.2	207991	2	BX545855	BX545855 Danio rer
10	21.8	87.2	267888	1	AP002565	AP002565 Escherich
11	20.8	83.2	71661	10	AL954643	AL954643 Mouse DNA
12	20.8	83.2	96879	9	HS384P21	AL022171 Human DNA
13	20.8	83.2	112484	10	AL954640	AL954640 Mouse DNA
14	20.8	83.2	151950	2	AC084813	AC084813 Homo sapi
15	20.2	80.8	2000	6	AX508952	AX508952 Sequence
16	20.2	80.8	5641	6	AX346299	AX346299 Sequence
17	20.2	80.8	89779	8	AB005234	AB005234 Arabidops
18	20.2	80.8	101371	8	AC002534	AC002534 Arabidops
19	20.2	80.8	145712	5	AL935136	AL935136 Zebrafish
20	20.2	80.8	182803	2	AC145899	AC145899 Pan trogl
21	20.2	80.8	186981	9	AC008459	AC008459 Homo sapi
22	20.2	80.8	216010	2	AC119517	AC119517 Rattus no
23	20.2	80.8	219476	9	AC068945	AC068945 Homo sapi
24	20.2	80.8	220414	2	BX546482	BX546482 Danio rer
25	20.2	80.8	240663	2	AC097541	AC097541 Rattus no
26	19.8	79.2	884	11	CNS06KTH	AL403338 T3 end of
27	19.8	79.2	57902	2	AC103790	AC103790 Homo sapi
28	19.8	79.2	68181	2	AC103972	AC103972 Homo sapi
29	19.8	79.2	94609	8	AB046439	AB046439 Arabidops
30	19.8	79.2	112615	9	AC079617	AC079617 Homo sapi
31	19.8	79.2	115626	9	AC103792	AC103792 Homo sapi
32	19.8	79.2	116461	8	AC051625	AC051625 Genomic S
33	19.8	79.2	121038	8	AP003342	AP003342 Oryza sat
34	19.8	79.2	121501	8	AC069557	AC069557 Genomic S
35	19.8	79.2	149098	8	AP003296	AP003296 Oryza sat
36	19.8	79.2	152813	2	AP005860	AP005860 Oryza sat
37	19.8	79.2	157295	2	AC079055	AC079055 Homo sapi
38	19.8	79.2	159302	2	AC098585	AC098585 Homo sapi
39	19.8	79.2	160295	10	AL845504	AL845504 Mouse DNA
40	19.8	79.2	160586	2	AC126319	AC126319 Mus muscu
41	19.8	79.2	178959	9	AC028402	AC028402 Homo sapi
42	19.8	79.2	182554	2	AC138307	AC138307 Mus muscu
43	19.8	79.2	183629	10	AC132596	AC132596 Mus muscu
44	19.8	79.2	183245	2	AC068090	AC068090 Homo sapi
45	19.8	79.2	195494	2	BX545912	BX545912 Danio rer

ALIGNMENTS

RESULT 1
ECARSRBC
LOCUS ECARSRBC 3492 bp DNA linear BCT 20-JUL-1995
DEFINITION E.coli genes arsr, arsb, arsc.
ACCESSION X80057
VERSION X80057.1 GI:510824
KEYWORDS arsb gene; arsc gene; arsenate reductase; arsenic-efflux pump;
arsenic-inducible repressor; arsr gene.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Diorio,C., Cai,J., Marmor,J., Shinder,R. and DuBow,M.S.
TITLE An Escherichia coli chromosomal ars operon homolog is functional in

Pred. No. is the number of results predicted by chance to have a

```

arsenic detoxification and is conserved in gram-negative bacteria
J. Bacteriol. 177 (8), 2050-2056 (1995)
MEDLINE 95238276
PUBMED 7721697
REFERENCE 2 (bases 1 to 3492)
AUTHORS Diorio,C.
TITLE Direct Submission
JOURNAL Submitted (06-JUL-1994) C. Diorio, McGill University, 3775
University Street, Montreal, Quebec, H3A 2B4, CANADA
FEATURES
    Location/Qualifiers
        1..3492
            /organism="Escherichia coli"
            /mol_type="genomic DNA"
            /strain="K12"
            /db_xref="taxon:562"
repeat_region 528..538
            /rpt_type=INVERTED
repeat_region 544..554
            /rpt_type=INVERTED
gene 655..1061
    /gene="arsR"
-35_signal 655..660
    /gene="arsR"
-10_signal 678..683
    /gene="arsR"
RBS 697..701
    /gene="arsR"
CDS 708..1061
    /gene="arsR"
    /codon_start=1
    /transl_table=1
    /product="arsenic-inducible repressor"
    /protein_id="CAA56361.1"
    /db_xref="GI:510825"
    /db_xref="GOA:P37309"
    /db_xref="SWISS-PROT:P37309"
    /translation="MSFLPIQLFKILADETRGLVLLLSLSELGELCVCDLCTALDQSQ
PKIHHALLRESGLLRLKQKWHYRLSPHAWAAKIIDEAWRCQEKVQIVRN
LARQCSGDSKNKICS"
gene 1071..2404
    /gene="arsB"
-35_signal 1071..1076
    /gene="arsB"
repeat_region 1072..1084
    /rpt_type=INVERTED
repeat_region 1087..1099
    /rpt_type=INVERTED
-10_signal 1088..1093
    /gene="arsB"
RBS 1106..1110
    /gene="arsB"
CDS 1115..2404
    /gene="arsB"
    /codon_start=1
    /transl_table=1
    /product="membrane-located arsenite efflux pump"
    /protein_id="CAA56362.1"
    /db_xref="GI:516211"
    /db_xref="GOA:P37310"
    /db_xref="SWISS-PROT:P37310"
    /translation="MLAGAFVLTVLIVWPKGLIGWGSATLGAVALVTGVWHPG
DIPVNVNNAATAPAFIVIIISLLDSGPFEMAAHVRWNGRGLLFTTWVLG
AAVALPANDGAAALITPIVTAMILALGSGKTTLAFVMAAGFIADTASLPIVSNLV
NIVSDPFGLGFRFASVWVPVDIAIVATLVMLHLYFRKIIPQYDNALLKSPAEAI
KDPATFKGVVLLLLVGFVPLEPGIPVSAIAVAGAILIFVVAARGHAINTKVLR
GAPQIVIFSLGMIVVYGLRNAGLTYSLVINADNGLWAAATLGTGFTALFSSI
MNMPTVLVGLSIDGTSASGVKEAMVYANVIGDLGPKITPIGSLATLLWLHLVLSQ
KMTTSWGYPRTGIIMTLPLVLFVTLAALALRLSPTL"
gene 2417..2842
    /gene="arsC"
CDS 2417..2842
    /gene="arsC"
    /codon_start=1

```

```

/transl_table=11
/product="arsenate reductase"
/protein_id="CAA56363.1"
/db_xref="GI:516212"
/db_xref="GOA:P37311"
/db_xref="SWISS-PROT:P37311"
/translation="MSNTIYHNPACGTSRNTLEMRNSGTEPIIHYLETPTPTDEL
VKLIADMGISVRALLRNKVPBELGLABDKFTDDRLLDFMLQHPILINRPIVVTPLG
TRLCRPSEVLEILPDAQKGAFSKDEKGVDSAGKSLK"
repeat_region 2892..2905
    /rpt_type=INVERTED
repeat_region 2910..2923
    /rpt_type=INVERTED
ORIGIN
Query Match 100.0%; Score 25; DB 1; Length 3492;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTAAGTCATATATGTTTTCACCTTA 25
Db 638 TTAAGTCATATATGTTTTCACCTTA 662
RESULT 2
AE000426 10240 bp DNA linear BCT 01-DEC-2000
LOCUS Escherichia coli K12 MG1655 section 316 of 400 of the complete
DEFINITION genome.
ACCESSION AE000426 U00096
VERSION AE000426.1 GI:1789910
KEYWORDS Escherichia coli K12
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
REFERENCE 1 (bases 1 to 10240)
AUTHORS Blattner,P.R.; Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
Mau,B. and Shao,Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)
97426617
9728503
2 (bases 1 to 10240)
Blattner,P.R.
Direct Submission
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
3 (bases 1 to 10240)
Blattner,P.R.
Direct Submission
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
4 (bases 1 to 10240)
Plunkett,G. III.
Direct Submission
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using GeneMark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology Atlanta, GA,
30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG

```

Site Nos., unique ID nos. for the genes in the *E. coli* Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the *E. coli* Genome Project's World wide Web site (<http://www.genetics.wisc.edu>). *** The *E. coli* K12 sequence and its annotations are periodically updated; this is version MS4. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

```

FEATURES
    source
        Location/Qualifiers
            1..10240
                /organism="Escherichia coli K12"
                /mol_type="genomic DNA"
                /strain="K12"
                /sub_strain="MG1655"
                /db_xref="taxon:83333"
            <1..31
                /notes="factor Sigma70; predicted +1 start at 3638369"
            161..1630
                /gene="yhiP"
                /note="synonym: b3496"
            161..1630
                /gene="yhiP"
                /function="putative transport; Not classified"
                /notes="0489; 99 pct identical amino acid sequence and
                equal length to YHIp_ECOLI SW: P36837"
                /codon_start=1
                /transl_table=11
                /product="putative transport protein"
                /protein_id="AAC76521.1"
                /db_xref="GI:1789911"
                /translation="MNITTPMGKLOQRPFFMIFVFWELWERFGYVQGVLAFFVKQ
                LQFSQQAFTFCFAALVYGLISIGYVGDHLGTRKIVLGAALVIAFTWGMGL
                KPDPLFALGTAVNGNLFKAPASLISKVCPKDPRLDGAFTLPYMSINIGSLIAL
                SLAPVADRGYSVTYVLAGAGLI IALLVVIACGMVKDGSPPDPFMSFKLLAYL
                LGSVMIFVCAMLMENVEANVLVLIVLIVIIIPRQAPKLDKTRKRMKVFVLM
                EAVFYLLAQMPSTLNFAINNVHHEILGFSINPVFQALNPFVWLASPILAGIYT
                HLGNKGLSMFMKFTLGMFCMSLGLTAAAGMWFADAGLTSPMFIIVLYLQSLG
                ELPISALGLAMIAALVPOHLMSEFLGMWFLTQAAFLGLGYVATFTAVPDNITDPLET
                LPVYTVFGKIGLVTLGVAVMMLMVPWPKRMIAFPESH"
                complement (1679..2431)
                /gene="yhiQ"
                /note="synonym: b3497"
                complement (1679..2431)
                /gene="yhiQ"
                /function="orf; Unknown"
                /notes="f250; 100 pct identical to 218 amino acids of
                YHIQ_ECOLI SW: P37633 but has 32 additional N-ter
                residues; gtg start, alternate starts possible"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC76522.1"
                /db_xref="GI:1789912"
                /translation="MKICIDETGTGGLSVLAARWGLEHDEHDLNWLVLTPHELEL
                RKDEPKLGSIFVDFVGGMAHPRKFGGGEAVAKVGIKGDYLPDVYDATGLGRD
                AFVLASVGRMLERNFVVALLDGLARGYADAIEIGWLQRLOLIHASSITALTD
                ITPRQVYLDPMFPFHQSALVKYKEMVFPQSLVGLDPLDGLLEPARLLATKRVVVK
                RPDYAPPLANVATPNVATVTKGRFDIYAGTPV"
                complement (2439..4481)
                /gene="prlC"
                /note="synonym: b3498"
                complement (2439..4481)
                /gene="prlC"

```

```

/EC_number="3.4.24.70"
/function="enzyme; Degradation of proteins, peptides,
glyco"
/notes="f680; 100 pct identical to OPDA_ECOLI SW: P27298;
CG Site No. 18031"
/codon_start=1
/transl_table=11
/product="oligopeptidase A"
/protein_id="AAC76523.1"
/db_xref="GI:1789913"
/translation="MTNPLLPFFLPSPSKILPHEVPAVTKALNDCRENVVVAAQ
APYTWENLCPLAEVDVILGFSPVSHLNVKNSPELREAYEQTLPLISEYTWVQG
HEGLYKAYRDLDGDHYATLNATQKAVDNALRDFELSGIGLKPKEKQYGEIATRLS
BLGNOYSNNVLDTAGMTKLVTDKAEALAGMPESALAAKAQAEKELGEGYLLITLIPS
LIPVMTYCDNALRESEYRAYSTRASDQGNAGKWDNSKVMEEILALRHELAQLGFFE
NYAKSLATMAENPQOVLDTLAKARQPKGEKLAQLAPAKARFGVDLOPMDI
AYSEKQHLIYSIDEOIRYPFENKAYNGLFEVKRIYGITAKERQDVDMHEDVR
FPEYDENNELKRGSYLILYAKENKRGAMWDDCVGQMKADGSLQKFAVLTCTNFR
PVNGKPAFTHDEVITLPHFEGHGLHMLTITAGVSGISGVPMADVAVELSPQFNW
CWEPKALAFISGHYETGPELPEKLLDKMLAAANYQAAFLFILRQLEFGFLDFRLHAAR
PDQAKILTELAIEIKLVAVVPSWGRFPHAFSHIFAGGYAGYYSYLVADVLAAADA
FSRPEEGIFNRETGQSFLDNILSRGSEEPMDLFRPRGEPQDLAMLEHYGK"
complement (2457..2488)
/gene="prlC"
/notes="factor Sigma70; predicted +1 start at 3640781"
complement (2545..2575)
/gene="prlC"
/notes="factor Sigma70; predicted +1 start at 3640869"
complement (2580..2608)
/gene="prlC"
/notes="factor Sigma70; predicted +1 start at 3640904"
complement (4546..4573)
/notes="factor Sigma32; predicted +1 start at 3642870"
4561..4587
/notes="factor Sigma70; predicted +1 start at 3642925"
4684..5526
/gene="yhiR"
/notes="synonym: b3499"
4684..5526
/gene="yhiR"
/function="orf; Unknown"
/notes="0280; 100 pct identical amino acid sequence and
equal length to YHIR_ECOLI SW: P37634"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC76524.1"
/db_xref="GI:1789914"
/translation="MLSYRHSFHAGNHADVLKHTVQSLLIESLKEKDKPFLYLDTHAG
AGRYOLGSEHAERTGEYLEGIARIWOODLPAELRAYINVKHFNRSGLRYPGSPPL
IARLLRSQDSIQLTSLHPSDYPLRSFQDSRVRKADQFQOLKAKLPVSRGL
ILLDPPEKTDYQAVSGIAEGYKFAFGIYALVPVLRQIQIKEMHDLLEATGIRK
ILQIELAVLPDSDRKMTASGMINVINPWKLEQQMNVLPWLHSLKLVPAAGTGHATVSW
IVPE"
5291..5303
/gene="yhiR"
/notes="central position to predicted promoter: -157.5"
/bound_moiety="Ada predicted site"
5421..5449
/gene="yhiR"
/notes="factor Sigma70; predicted +1 start at 3643787"
5598..6950
/gene="gor"
/notes="synonym: b3500"
5598..6950
/gene="gor"
/EC_number="1.6.4.2"
/function="enzyme; Biosynthesis of cofactors, carriers:
Thioredoxin, glutaredoxin, glutathione"
/notes="0450; 100 pct identical to GSHR_ECOLI SW: P06715;
CG Site No. 665"
/codon_start=1
/transl_table=11

```



```

/product="glutathione oxidoreductase"
/protein_id="AAC76525.1"
/db_xref="GI:1789915"
/translation="MTKYDYDIAIGGSGGIASINRAAMVQKCALTEAKELGCTCVN
VGVCPKVMWAAQIREAHMVGPDYGFDTINKFNWETLIASRTAYIDRTHTSYVN
LGNMVDVKGARFVDKATLEVNGETITADHILIATGGRPSHPIQGVEXGIDSGF
PALPALPERVAVVAGYIAVELAGVINGLAKTHLFVRKHAPLRSFDFMISETLVEVM
"
Query Match      100.0%; Score 25; DB 1; Length 10240;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTATGTCATATATGTTTGTACTTA 25
|||||
Db 7757 TTAAGTCATATATGTTTGTACTTA 7781

```

```

RESULT 3
LOCUS      AB015361      11524 bp      DNA      linear      BCT 18-OCT-2002
DEFINITION Shigella flexneri 2a str. 301 section 324 of 412 of the complete
            genome.
ACCESSION  AB015361 AB005674
VERSION    AB015361.1 GI:24053992
KEYWORDS
SOURCE
ORGANISM   Shigella flexneri 2a str. 301
            Shigella flexneri 2a str. 301
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
            1 (bases 1 to 11524)
            Jin, Q., Yuan, Z.H., Xu, J.G., Wang, Y., Shen, Y., Lu, W.C., Wang, J.H.,
            Liu, H., Yang, J., Sun, L.L., Xue, Y., Zhang, X.B., Zhang, J.Y., Yang, G.W.,
            Wu, H.T., Dong, J., Sun, L.L., Xue, Y., Zhao, A.L., Gao, Y.S., Zhu, J.P.,
            Kan, B., Chen, S.X., Yao, Z.J., He, B.K., Chen, R.S., Ma, D.L.,
            Qiang, B.Q., Wen, Y.M., Hou, Y.D. and Yu, J.
            Genome sequence of Shigella flexneri 2a: insights into
            pathogenicity through comparison with genomes of Escherichia coli
            K12 and O157
            Nucleic Acids Res. 30 (20), 4432-4441 (2002)
            12384590
            2 (bases 1 to 11524)
            Jin, Q., Shen, Y., Wang, J.H., Liu, H., Yang, J., Yang, P., Zhang, X.B.,
            Zhang, J.Y., Yang, G.W., Wu, H.T., Dong, J., Sun, L.L., Xue, Y.,
            Zhao, A.L., Gao, Y.S., Zhu, J.P., Chen, S.X., Yao, Z.J., Wang, J.,
            Lu, W.C., Qiang, B.Q., Wen, Y.M. and Hou, Y.D.
            Direct Submission
            Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry
            of Public Health, 100 Yingxin Jie, Xuanwu Qu, Beijing 100052, P.R.
            China

```

```

FEATURES
source
    1..11524
        /organism="Shigella flexneri 2a str. 301"
        /mol_type="genomic DNA"
        /strain="301"
        /serotype="2a"
        /db_xref="taxon:198214"
        67..1419
            /gene="gor"
            /locus_tag="SF3531"
            /locus_tag="SF3531"
            /locus_tag="SF3531"
            /note="Residues 1 to 450 of 450 are 99 pct identical to
            residues 1 to 450 of a 450 aa protein from Escherichia
            coli O157:H7 ref: NP_312399.1"
            /codon_start=1
            /transl_table=11
            /product="glutathione oxidoreductase"
            /protein_id="AAC44987.1"
            /db_xref="GI:24053993"
            /translation="MTKYDYDIAIGGSGGIASINRAAMVQKCALTEAKELGCTCVN
            VGVCPKVMWAAQIREAHMVGPDYGFDTINKFNWETLIASRTAYIDRTHTSYVN
            LGNMVDVKGARFVDKATLEVNGETITADHILIATGGRPSHPIQGVEXGIDSGF

```

```

PALPALPERVAVVAGYIAVELAGVINGLAKTHLFVRKHAPLRSFDFMISETLVEVM
NAEQPOLHTNAIPKAVPKNADGSLTLEDETSVDCLIWAGREFPANDINLEAAG
VTNEKGYIVVDKYQNTINISYAVGNTGAVELTPVAVAGRRLSERLFPNKPDEHL
DYSNIPTVFSPHPPIGTVGLTEPQARQYGDQVKVYKSSFTAMVTAVTHRQPCRMK
LVCVGPBEKVIQTHGIGFGMDLQGLQFVALKMGATKKQPDNTVAIHPTAAEEFVTR
"
complement (2332..2727)
/gene="insB"
/locus_tag="SF3532"
/complement (2332..2727)
/gene="insB"
/locus_tag="SF3532"
/note="Residues 1 to 131 of 131 are 100 pct identical to
residues 37 to 167 of a 167 aa protein from Shigella
flexneri pir: T44475"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AAM44988.1"
/db_xref="GI:24053994"
/translation="MIVCAEMDRHMGVVGAKSRQRLFYAYDRIRRTVVAHVFGERTL
ATLERLLSLISAFEVVWMDGHPLEYSLKGLHVISKEYTORIERHNLNLQHLAR
LVRSLSFSKSVBLHDKVIGHYINIKHYQ"
complement (2754..3029)
/gene="insA"
/locus_tag="SF3533"
/complement (2754..3029)
/gene="insA"
/locus_tag="SF3533"
/note="Residues 1 to 91 of 91 are 98 pct identical to
residues 1 to 91 of a 91 aa protein from Shigella flexneri
ref: NP_085394.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AAM44989.1"
/db_xref="GI:24053995"
/translation="MASISIRPCSCATEGVVNGKSTAGHQRYLCSHCRTWQLQFT
YTASQPTQHQKIIDVAMNGVGCASARIMGVGLTVLRHLKNSGRSR"
3071..3616
/locus_tag="SF3534"
3071..3616
/locus_tag="SF3534"
/note="No database matches"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAM44990.1"
/db_xref="GI:24053996"
/translation="MAALPCVPSNCSKFLMKPVWASFYCLANWESVASIALPSIS
RSPRYPAIMHCCWAGYCWTSKVSQFITYRRIPQGRKILMKPGDVNRKFRQIS
ATWPKTVPETVTFVAVNLAKHINICTVLGSRHYVTRRYLCPDRHIGYLAERFR
HRLFCNARRGTGVYGRGSSG"
3588..4772
/gene="arsB"
/locus_tag="SF3535"
3588..4772
/gene="arsB"
/locus_tag="SF3535"
/note="Residues 1 to 394 of 394 are 100 pct identical to
residues 36 to 429 of a 429 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_290073.1"
/codon_start=1
/transl_table=11
/product="arsenical pump membrane protein"
/protein_id="AAM44991.1"
/db_xref="GI:24053997"
/translation="MTVGTVHFGDIPVWNVNNTAAFTAVIITISLLDESQFFWEA
ALFVSRWNGRGRLLFTWVILGAVALPANDGAALITPIVIAMLLALGFSKGTFL
AFVMAAGPIADTASLPLIVSNLVNSADPFGFPGFVYASVMVVDIAIVATLWMLH
LYRKDIPQNDMALLKSPARIKDPATFGVWVLLLLVLPFVLEPLGIPVSAIAA
VGLLILVFAKRGHINTGVKRCAPQIVIFSLGMVLYVYGLRNAGLTYLSCVLNV
LADNGLWAATLGTGFLTAFLSSNNNNPTVVLGALSIDGSTAGVKEAMVYANVIC

```

```
DLGPKITPIGSLATLLHLVLSQNMNTISNGYYFTGTIIINTLPVLVFTVLAALALRLSP
TL"
gene      4785..5210
/genes="arsC"
/locus tag="SF3536"
CDS       4785..5210
/genes="arsC"
/locus tag="SF3536"
/notes="Residues 1 to 141 of 141 are 100 pct identical to
coli K12 ref: NP_417960.1"
/codon start=1
/transl table=11
/product="arsenate reductase"
/protein_id="AA044992.1"
/db_xref="GI:24053998"
VKLIADMGISVRALLRNKVEPEELGLAEDKFTDDLIDFMLQHPILINRPIVPTPLG
TLRCRSEVLEILPDAQKGAFSKEDGKVDAGKELK"
5560..6063
/genes="insB"
/locus tag="SF3537"
CDS       5560..6063
/genes="insB"
/locus tag="SF3537"
/notes="Residues 1 to 167 of 167 are 100 pct identical to
flexneri pir: T44475"
/codon start=1
/transl table=11
/product="orf, conserved hypothetical protein"
/protein_id="AA044993.1"
/db_xref="GI:24053999"
/translation="MSKQCHYGRWPQHGFTSLKLRPOSVTSRIQPGSDVIVCAEMD
EMWYGVAKSRQWLYPYADRINTRVVAHVFGERTLATLRLLLSLSAFEVVMVMTDG
WPLYESRLKGLHVISKRYTORIERHNLRLQRLARLVKRSLSFSKSVLHDKVIGHY
LNIRYQ"
6088..7050
/locus tag="SF3538"
CDS       6088..7050
/locus tag="SF3538"
/notes="Residues 34 to 319 of 320 are 34 pct identical to
jejunii ref: NP_282688.1"
/codon start=1
/transl table=11
/product="purative membrane protein"
/protein_id="AA044994.1"
/db_xref="GI:24054000"
/translation="MCFVFLAVELSLFIVISAGVSLIRQKVPDKIQOMGARKGRGY
LIALILGAVTPFCSTIPMLRGLGAKAGFGPTLFLFVSLINPFIIVGLMMTFGW
KVTLLYAIAGVSVLASIILDSLGERHIIASKSSANCCAPAKTSPGTVTYPIKVS
CCSPAAKIENPVVTCSTKAVYSINPIKATDQKQKDVLPYLLGLVLSGPIYG
PIPSAMTAAHAGNDNPAPLPSAVGILYIIRAVIPLASVIMTKMGLGALMALII
GSAGSLTEVILKSMFRPMIAAFUIVILGMALMGYLTQLLP"
7631..8854
/locus tag="SF3539"
CDS       7631..8854
/locus tag="SF3539"
/notes="Residues 1 to 365 of 407 are 75 pct identical to
coli O15:H7 ref: NP_312403.1"
/codon start=1
/transl table=11
/product="orf, conserved hypothetical protein"
/protein_id="AA044995.1"
/db_xref="GI:24054001"
/translation="MSIDTPGIIINTYHGDLYNCTNTDNVKTPTPWPCEWEEQQ
PINSTSGYNEQDFDLAQQLQOINACHTNTYNNADYSKVVVOLSLINNIETIS
STPLTQOTSLINQHNIYREKDSVCQIIVVNPENNPITIKILVEEGIPERFSV
QTSSDNKQFAGRADLPDTRIDISLYLMAKLYEHSKNEKMEALAGCIDFNFM
TGDMSKLYLTSKPYFEDLLINFTDANLFTIPSHKENPIPLKHKYEQHLDKQNG
LPSTLLTINDNSLRAXAEIASRIIDPLEAKVNVLSPPDILKYKQFKICYQLQEFT
```

```
Query Match      100.0%; Score 25; DB 1; Length 11524;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTGGACTTA 25
   |||||
Db 2229 TTAAGTCATATATGTTTGGACTTA 2253
   |||||

RESULT 4
AC145934/c 179941 bp DNA linear HTG 01-AUG-2003
LOCUS      Gallus gallus chromosome UNK clone CH261-4211, *** SEQUENCING IN
DEFINITION PROGRESS ***, 44 unordered pieces.
AC145934
VERSION    AC145934.1 GI:33386884
KEYWORDS   HTG; HTGS; PHASE1
SOURCE     Gallus gallus (chicken)
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE  1 (bases 1 to 179941)
AUTHORS   Wilson,R.K.
TITLE     The sequence of Gallus gallus clone
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 179941)
AUTHORS   Wilson,R.K.
TITLE     Direct Submission
JOURNAL   Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submission@wustl.edu
----- Project Information -----
Center project name: J_AA042101

* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1176: contig of 1176 bp in length
* 1177 1276: gap of unknown length
* 1278 2558: contig of 1282 bp in length
* 2559 2659: gap of unknown length
* 2659 3488: contig of 1190 bp in length
* 3488 3948: gap of unknown length
* 3949 5267: contig of 1318 bp in length
* 5267 5366: gap of unknown length
* 5366 6401: contig of 1035 bp in length
* 6401 6501: gap of unknown length
* 6501 7599: contig of 1098 bp in length
* 7599 7600: gap of unknown length
* 7600 8832: contig of 1132 bp in length
* 8832 10481: contig of 1550 bp in length
* 10481 10581: gap of unknown length
* 10581 12037: contig of 1456 bp in length
* 12037 12137: gap of unknown length
* 12137 13596: contig of 1459 bp in length
* 13596 14911: gap of unknown length
* 14911 15011: contig of 1215 bp in length
* 15011 16379: contig of 1268 bp in length
* 16379 16379: gap of unknown length
```

```
* 16380 17948: contig of 1569 bp in length
* 17949 18048: gap of unknown length
* 18049 19612: contig of 1564 bp in length
* 19613 19712: gap of unknown length
* 19713 21723: contig of 2011 bp in length
* 21724 21824: gap of unknown length
* 21824 25266: contig of 3443 bp in length
* 25267 25366: gap of unknown length
* 25367 27105: contig of 1739 bp in length
* 27106 27205: gap of unknown length
* 27206 28536: contig of 1331 bp in length
* 28537 28637: gap of unknown length
* 28637 30905: contig of 2269 bp in length
* 30906 31005: gap of unknown length
* 31006 33224: contig of 2219 bp in length
* 33225 33324: gap of unknown length
* 33325 37384: contig of 4060 bp in length
* 37385 37484: gap of unknown length
* 37485 40797: contig of 3313 bp in length
* 40798 40897: gap of unknown length
* 40898 43117: contig of 2220 bp in length
* 43118 43217: gap of unknown length
* 43218 46243: contig of 3026 bp in length
* 46244 46343: gap of unknown length
* 46344 49765: contig of 3422 bp in length
* 49766 49865: gap of unknown length
* 49866 56038: contig of 6173 bp in length
* 56039 56138: gap of unknown length
* 56139 60188: contig of 4050 bp in length
* 60189 60288: gap of unknown length
* 60289 67106: contig of 6718 bp in length
* 67107 68978: contig of 1772 bp in length
* 68979 72005: gap of unknown length
* 72006 72005: contig of 3927 bp in length
* 72006 77513: gap of unknown length
* 77514 77613: contig of 4508 bp in length
* 77614 81177: contig of 3564 bp in length
* 81178 86112: gap of unknown length
* 86113 86712: contig of 5335 bp in length
* 86713 91513: contig of 4801 bp in length
* 91514 91614: gap of unknown length
* 91614 97537: contig of 5924 bp in length
* 97538 103313: contig of 5675 bp in length
* 103313 111607: contig of 8195 bp in length
* 111608 123017: gap of unknown length
* 123017 123116: contig of 11309 bp in length
* 123117 131760: contig of 8644 bp in length
* 131761 133264: contig of 1404 bp in length
* 133265 133364: gap of unknown length
* 133365 140148: contig of 6784 bp in length
* 140149 152045: contig of 11797 bp in length
* 152046 152145: gap of unknown length
* 152146 165190: contig of 13045 bp in length
* 165191 179941: gap of unknown length
* 165291 179941: contig of 14651 bp in length.
```

FEATURES

```
source
1..179941
/organism="Gallus gallus"
/mol_type="genomic DNA"
/db_xref="taxon:9031"
/chromosome="UNK"
/clone="CH261-4211"
1..1176
/note="assembly_name:Contig10"
```

misc_feature

misc_feature

```
misc_feature
/note="assembly_name:Contig17"
2659..3848
/note="assembly_name:Contig25"
3949..5266
/note="assembly_name:Contig30"
5367..6401
/note="assembly_name:Contig31"
6502..7539
/note="assembly_name:Contig32"
7700..8831
/note="assembly_name:Contig36"
8932..10481
/note="assembly_name:Contig39"
10582..12037
/note="assembly_name:Contig40"
12138..13596
/note="assembly_name:Contig41"
13697..14911
/note="assembly_name:Contig42"
15012..16279
/note="assembly_name:Contig43"
16380..17948
/note="assembly_name:Contig44"
18049..19612
/note="assembly_name:Contig45"
19713..21723
/note="assembly_name:Contig46"
21824..25266
/note="assembly_name:Contig47"
25367..27105
/note="assembly_name:Contig48"
27206..28536
/note="assembly_name:Contig49"
28637..30905
/note="assembly_name:Contig50"
31006..33224
/note="assembly_name:Contig51"
33325..37384
/note="assembly_name:Contig52"
37485..40797
/note="assembly_name:Contig53"
40898..43117
/note="assembly_name:Contig54"
43218..46243
/note="assembly_name:Contig55"
46344..49765
/note="assembly_name:Contig56"
49866..56038
/note="assembly_name:Contig57"
56139..60188
/note="assembly_name:Contig58"
60289..67006
/note="assembly_name:Contig59"
67107..68878
/note="assembly_name:Contig6"
68979..72905
/note="assembly_name:Contig60"
73006..77513
/note="assembly_name:Contig61"
77614..81177
/note="assembly_name:Contig62"
81278..86612
/note="assembly_name:Contig63"
86713..91513
/note="assembly_name:Contig64"
91614..97537
```

```
Query Match 100.0%; Score 25; DB 2; Length 179941;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 TTAACTCATATATGTTTTCACCTTA 25

|||||

```
Db 132284 TTAAGTCATATATGTTTGTACTTA 132260

RESULT 5
ECOUW76
LOCUS 225419 bp DNA linear BCT 07-NOV-1996
DEFINITION E. coli chromosomal region from 76.0 to 81.5 minutes.
ACCESSION U00039
VERSION U00039.1 GI:466582
KEYWORDS
SOURCE
  Escherichia coli
  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
  Enterobacteriaceae; Escherichia.
REFERENCE
  1 (bases 1 to 225419)
  Sofia,H.J., Burland,V., Daniels,D.L., Plunkett,G. III and
  Blattner,F.R.
  Analysis of the Escherichia coli genome. V. DNA sequence of the
  region from 76.0 to 81.5 minutes
  Nucleic Acids Res. 22 (13), 2576-2586 (1994)
JOURNAL
MEDLINE 94316500
CURED 8041620
REFERENCE
  2 (bases 1 to 225419)
  Plunkett,G. III.
  Direct Submission
  Submitted (25-MAR-1994) Guy Plunkett III, Laboratory of Genetics,
  University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
  Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
  608-263-7459
COMMENT
  This sequence was determined as part of the E. coli Genome Project
  (Frederick R. Blattner, director) at the University of
  Wisconsin-Madison. Supported by award HG00301 from the NIH Human
  Genome Project. The entire sequence was independently determined
  from E. coli MG1655; overlaps and conflicts with other sequence
  determinations are annotated. The start of this entry overlaps the
  start of the entry ECOWR82 (U10328) by 547 bp.
FEATURES
  source
    1..225419
    /organism="Escherichia coli"
    /mol_type="genomic DNA"
    /strain="K-12"
    /sub_strain="MG1655"
    /db_xref="taxon:562"
    /notes="This sequence comprises the following lambda
    clones: EC19-98, EC27-273, EC17-131, EC30MM5, EC30K509-3,
    EC27-334, EC30K606A-14, EC30K606A-12, EC30-154, EC27-385,
    EC27-1129, EC30MM9, EC27-239, EC30MM3, EC27-SF3955,
    EC18-200, EC27-1135, EC30R3900-8B, EC30R3900-5B, EC19MM6,
    EC30K573-3; M13mp19 or Janus vectors were used for
    subcloning"
  misc_feature
    complement (<1..1221)
    /note="corresponds to M28722; EC0GGT (928..2148)"
  gene
    complement (<1..1085)
    /gene="ggt"
    /notes="ggt"
    /BC_number="2.3.2.2"
    /note="CG Site No. 18280; cold-regulated"
  CDS
    complement (<1..1085)
    /gene="ggt"
    /transl_table=1
    /product="gamma-glutamyltranspeptidase"
    /protein_id="AAB18422.1"
    /db_xref="GI:466583"
    /translation="MIKPTFLARVAIAALLSGSCFSAARAAPPVSVYGVREDVFPV
    RAKQNVASVDATATQGVVDILKEGNVADAAVAGYALAVTHPOAGNLGGGGFMLIR
    SKNGNTAIDFREPAPAKATRMFLDDQNPDSKKSLTSHLSAGTPTGTVAGPSLADK
    YGTMLNKVQVPAFLKARDGFIYNDALADLLKYTGSEVLPNPHNSKALFWKEGEPLKK
    GTLVQANLAKLEIAENGPDPEFYGTIAEQIAEQMGKNGGLITKEDLAAYKAVERT
    PISGDVGRGVQVSMPPSPSGGIHIVQILNILENFDMMKYFGSADAMQIMAAEKAY
    ADSREVLGPDFFVKVPEWQALTNKAYKS"
  misc_feature
    <60..35171
    /note="corresponds to lambda clone EC19-98"
  promoter
    1067..1094
```

```
/note="39%"
complement (1085..1112)
/note="43%"
complement (1146..1173)
/note="45%"
complement (1181..2402)
/note="corresponds to X56908; ECUGPQQ (1..1222)"
1205..1645
/note="alternate gene name yhhA"
/codon_start=1
/transl_table=11
/label=ORF_0146a
/protein_id="AAB18423.1"
/db_xref="GI:466584"
/translation="MKRIILITALLPVFGFAQPIINTLANPNOGQCIQSOQMOTOMQ
TQOIQKXMLNQOLKTYTQLOQQLHLENNNSORVLOSQGERNPASQQLPNTNGG
MLNSNRNPDSLSNQHHMPERRNGMDLNPSTPQPDIFKTIKP"
complement (1565..2383)
/note="corresponds to X14437; ECUGPQ (1..819)"
gene
  complement (1632..2375)
  /gene="ugpQ"
  complement (1632..2375)
  /gene="ugpQ"
  /EC_number="3.1.4.2"
  /codon_start=1
  /transl_table=11
  /product="glycerophosphocholine phosphodiesterase"
  /protein_id="AAB18424.1"
  /db_xref="GI:466585"
  /translation="MSNWPYRIYAHRCGGKLAPENTLASIDVGAKYGHKMLEPDAKL
  SKDGIFLHDNLTERTNGVAGELNWLQDLRVDSGWSKMPKGPPLISQVAE
  KCRHEGMANIEIKPTTGTLGTGVAALAEALWAGMTPLLSSEFIDALAAQQA
  PELPGLLDEDRDDRELTALRGCVSHLNKLNKARVMQLKDAGLRILVYTVNKP
  QRAELLNGVDCICTDAIDVIGNFTAQ"
  1693..1727
  /note="terminator-like sequence"
  2192..2323
  /note="predicted bend of 73.99 degrees"
  complement (2225..16941)
  /note="corresponds to X13141; ECUGP (1..4717)"
  gene
    complement (2372..3481)
    /gene="ugpC"
    complement (2372..3481)
    /gene="ugpC"
    /function="sn-Glycerol-3-phosphate transport system"
    /note="CG Site No. 17794"
    /codon_start=1
    /transl_table=11
    /protein_id="AAB18425.1"
    /db_xref="GI:912455"
    /translation="MPSCAAMSIYRNKAGLKLAQVTKSDGKTQVIKPLTLDVADGE
    FIYVVGSCGCKSTILRNVAQLERVTEGDIWINDQVTEMEPKDGLAMVFQNYALYP
    HNSVEHNAWGLKIRMGKQKQIAERKVAARILELDGLKRPRLSGGQQRQVAMGR
    AIVRDPAYFLDPLSLNDKLRVQMLRLEQLQHLHRLKTSLYVTHDQVEAMTLAQRV
    MNGGVAAQIGTPEYVEKPSLFVAFSGPAMNLTGRVNEGTFFELDGGTFLP
    LVNGYRVAGRKMTIGIRPEHIALSSQAAGGVPMVMDTLEILGADNLAHGRWGEQCLV
    VRLAHERPTAGSTLWLHLAENQLHLFDGETGQRV"
    complement (3444..4289)
    /gene="ugpB"
    complement (3444..4289)
    /gene="ugpB"
    /function="sn-Glycerol-3-phosphate transport system"
    /note="CG Site No. 17791; membrane protein"
    /codon_start=1
    /transl_table=11
    /protein_id="AAB18426.1"
    /db_xref="GI:466587"
    /translation="MIENRPLWTFISHTMLILGIADVILFPLYVFAVATLDKQAVAA
    FMTPTGTHLENIHNWVNGTNSAPFWMLNLSFWMAFSITLGIKTVSKLSAFAI
    PWTFPPLNLFPMWIFITLMLPVEVRIPPTVEVIANLQMLDSYAGLTFLPMASATPF
    LFRPFTLPDELVEAARDIGASPMRPFCDIVPLSKTNLAALPVITTYGNQVILWP
    LLITFDVLGTIVAGIKMIATGCTTIEWNSVMVAMLLTLPVPVILVMQRAFVRGL
    VDSEK"
```

gene complement(4286..5173)
/gene="ugpa"
CDS complement(4286..5173)
/gene="ugpa"
/function="sn-glycerol-3-phosphate transport system"
/note="CG site No. 40"
/codon_start=1
/transl_table=11
/protein_id="AAB18427.1"
/db_xref="GI:466598"
/translation="MSSRRPVRSLWPLLVAPOLLITVIFFTWPAEALWYLSQSV
DPFGSSQFVLGNDVTLFHDSEYILDSFTTKFTFTVVSGLLVSLFFAALVYIVR
GSRFYQTLMLPYAVAPAAVLMIFLFPNGRGLITFLAEFGYDWNHQAQSGAMEL
VVFASVWQISYNYFLPFAALQISPRSLIEAAIDGAGFIRPFKIALPLIAPVSPFL
LVNLVYAFFTDFFVIDAATSGQPVQAATLTLYKIYRGGFTGLDASSAQSVMVLMFL
VIVLTVOFRYVESKVRVQ"
gene complement(5271..6587)
/gene="ugp8"
CDS complement(5271..6587)
/gene="ugp8"
/function="sn-glycerol-3-phosphate transport system"
/note="CG site No. 39; periplasmic binding protein"
/codon_start=1
/transl_table=11
/protein_id="AAB18428.1"
/db_xref="GI:466599"
/translation="MKPLHYTASALALGLALMGNAQAVTTIPFVHSMGEGLGKVDLSL
AQRNAPDVKIVYPTVGNVQNSAGIAAIFRTGNAPAILQVYEVGTATWASKAKI
PYDVFKAQIGDSQVPTVSGYSDSKTGLHSQPFNSSTPLYTNKDAFKKAGL
DPEQPKTWQDLADYAKASGKCGYASGQWIOLENFSAWNGLPFASKNGNFGDG
TDVLEFNKPEOVHIALEEMNKKGDSYVGRKDETEKFGVNGCAMTTASSGSLAN
IREYAKFNGVGMVDPDADAKAPONAIIIGASLWVMQGDKEYTVGVAKFLDPLAKP
ENAAEHQKTYGLPITKAAVLDLTREQGFYENPGADTATROMLNKPPFPFTKGLRLGN
MPQIRVIVDEELSVWTGKTPQQAOLDTAVRGNQLRRFRKSTKS"
misc_difference 5473..5474
/note="GC in X13141; SS here"
misc_difference 5588
/note="T in X13141; G here"
misc_difference 6021..6024
/note="GGCG in X13141; SSGS here"
misc_feature complement(6618..15331)
/note="corresponds to J05516; ECOLIVHMGF(1..8703)"
6640
/note="mRNA start determined by S1 mapping; Molecular
Microbiol. 2:767 (1988)"
promoter evidence=experimental
complement(6651..6679)
/note="includes pho box"
promoter complement(6695..6723)
/note="768"
repeat_region 6826..6859
/standard_name="RBP; repetitive extragenic palindromic
element"
/note="contains 1 REP sequence"
misc_difference 6836..6837
/note="CG in J05516; GC in X13141 and here"
misc_difference 6845..6846
/note="C in J05516; CC in X13141 and here"
gene complement(6986..7711)
Query Match 100.0%; Score 25; DB 1; Length 225419;
Best Local Similarity 100.0%; Pred. NO. 6.1;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTAAGTCATATATGTTTGGACTTA 25
|||||
Db 62717 TTAAGTCATATATGTTTGGACTTA 62741
|||||
RESULT 6
AC146183 242495 bp DNA linear HTG 04-NOV-2003
LOCUS
DEFINITION Pan troglodytes chromosome y clone CH251-548L16, *** SEQUENCING IN

PROGRESS ***, 6 unordered pieces.
AC146183
VERSION AC146183.2 GI:38154191
KEYWORDS HTG; HTGS_PRAESI.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
1 (bases 1 to 242495)
Wilson, R.K.
The sequence of Pan troglodytes clone
Unpublished
2 (bases 1 to 242495)
Wilson, R.K.
Direct Submission
Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 242495)
Wilson, R.K.
Direct Submission
Submitted (04-NOV-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
On Nov 4, 2003 this sequence version replaced gi:33387136.
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: bases at least Q40
Consensus quality: bases at least Q30
Consensus quality: bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1269: contig of 1269 bp in length
* 1270 1369: gap of unknown length
* 1370 2935: contig of 1566 bp in length
* 2936 3035: gap of unknown length
* 3036 4616: contig of 1581 bp in length
* 4617 4718: gap of unknown length
* 4719 6291: contig of 1575 bp in length
* 6292 8255: gap of unknown length
* 8256 8354: contig of 1863 bp in length
* 8355 242495: contig of 234141 bp in length.
FEATURES
Location/Qualifiers
1. 242495
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="y"
/clone="CH251-548L16"
1..1269
/note="assembly_name:Contig12"
1370..2935
/note="assembly_name:Contig13"
3036..4616
/note="assembly_name:Contig14"
4717..6291
/note="assembly_name:Contig18"
6392..8254
/note="assembly_name:Contig19"
8355..242495
/note="assembly_name:Contig20"
ORIGIN

Query Match 100.0%; Score 25; DB 2; Length 242495;
 Best Local Similarity 100.0%; Pred. No. 6;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTGGACTTA 25
 |||||
 Db 3329 TTAAGTCATATATGTTTGGACTTA 3353

RESULT 7
 LOCUS AE016992/c 289816 bp DNA linear BCT 22-APR-2003
 DEFINITION Shigella flexneri 2a str. 2457T section 15 of 16 of the complete genome

ACCESSION AE016992 AE014073
 VERSION AE016992.1 GI:30043426

KEYWORDS Shigella flexneri 2a str. 2457T
 SOURCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.

ORGANISM Shigella flexneri 2a str. 2457T
 REFERENCE Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W., Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A., Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S., Schwartz, D.C. and Blattner, F.R.
 Complete Genome Sequence and Comparative Genomics of Shigella flexneri Serotype 2a Strain 2457T
 Infect. Immun. 71 {5}, 2775-2786 (2003)

TITLE Shigella flexneri 2a Strain 2457T
 JOURNAL 12704152
 PUBMED 2 (bases 1 to 289816)

REFERENCE Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W., Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A., Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S., Schwartz, D.C. and Blattner, F.R.
 Direct Submission
 Submitted (13-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA

TITLE Location/Qualifiers
 JOURNAL 1. 289816
 REFERENCE /organism="Shigella flexneri 2a str. 2457T"

FEATURES
 source /mol_type="genomic DNA"
 /strain="2457T"
 /serotype="2a"
 /db_xref="taxon:198215"
 complement(238. .450)
 /gene="cspA"
 /locus_tag="S4179"
 complement(238. .450)
 /gene="cspA"

gene /locus_tag="S4179"
 /function="regulator; Adaptations, atypical conditions"
 /note="residues 1 to 70 of 70 are 100.00 pct identical to residues 1 to 70 from Escherichia coli K-12 : B3556"

CDS /codon_start=1
 /transl_table=11
 /product="cold shock protein 7.4, transcriptional activator of hns"
 /protein_id="AAP19147.1"
 /db_xref="GI:30043427"

gene /translation="MSGKMTGIVKWFNADKGFPTPDGSKDVPVHFSALQNDGYKS
 LDEGKVFTIESGKGPAGNVTSL"
 complement(896. .1021)
 /gene="yiaG"
 /locus_tag="S4180"
 complement(896. .1021)
 /gene="yiaG"

CDS /locus_tag="S4180"
 /function="regulator; Adaptations, atypical conditions"
 /note="residues 1 to 41 of 41 are 100.00 pct identical to residues 1 to 41 of 96 from Escherichia coli K-12 : B3555"

gene /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein"
 /protein_id="AAP19151.1"
 /db_xref="GI:30043431"

CDS /translation="MSGALAVSGCTTNPYTGEEAGKSAIGAGLSIVGAGIGALSSS
 KKDRGKALIGAAAGLGGVGYMDVQAKLDRKMGTVSVTRSGDNILINPNP
 VTPSSGAPLKPAGANTLTGVAVLVKYPKTVANVIGVYDSTGCHDLNRLSQORADS
 VASALITQGVDSRITQTGLGPANPIASNSTAECKAQRNRVEITLSPL"
 complement(896. .1021)
 /gene="yiaG"
 /locus_tag="S4180"
 complement(896. .1021)
 /gene="yiaG"

gene /locus_tag="S4180"
 /function="regulator; Adaptations, atypical conditions"
 /note="residues 1 to 41 of 41 are 100.00 pct identical to residues 1 to 41 of 96 from Escherichia coli K-12 : B3555"

CDS /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein"
 /protein_id="AAP19151.1"
 /db_xref="GI:30043431"

gene /translation="MSGALAVSGCTTNPYTGEEAGKSAIGAGLSIVGAGIGALSSS
 KKDRGKALIGAAAGLGGVGYMDVQAKLDRKMGTVSVTRSGDNILINPNP
 VTPSSGAPLKPAGANTLTGVAVLVKYPKTVANVIGVYDSTGCHDLNRLSQORADS
 VASALITQGVDSRITQTGLGPANPIASNSTAECKAQRNRVEITLSPL"
 complement(896. .1021)
 /gene="yiaG"
 /locus_tag="S4180"
 complement(896. .1021)
 /gene="yiaG"

CDS /locus_tag="S4180"
 /function="regulator; Adaptations, atypical conditions"
 /note="residues 1 to 41 of 41 are 100.00 pct identical to residues 1 to 41 of 96 from Escherichia coli K-12 : B3555"

gene /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein"
 /protein_id="AAP19151.1"
 /db_xref="GI:30043431"

CDS /translation="MSGALAVSGCTTNPYTGEEAGKSAIGAGLSIVGAGIGALSSS
 KKDRGKALIGAAAGLGGVGYMDVQAKLDRKMGTVSVTRSGDNILINPNP
 VTPSSGAPLKPAGANTLTGVAVLVKYPKTVANVIGVYDSTGCHDLNRLSQORADS
 VASALITQGVDSRITQTGLGPANPIASNSTAECKAQRNRVEITLSPL"
 complement(896. .1021)
 /gene="yiaG"
 /locus_tag="S4180"
 complement(896. .1021)
 /gene="yiaG"

/product="hypothetical protein"
 /protein_id="AAP19148.1"
 /db_xref="GI:30043428"
 /translation="MEYKDPHELLSLLEQIVFXDETQKIFLTHTTSCTEIEQL"
 1335. .2165
 /gene="yiaF"
 /locus_tag="S4181"
 1335. .2165
 /gene="yiaF"
 /locus_tag="S4181"
 1335. .2165
 /gene="yiaF"
 /locus_tag="S4181"
 /note="residues 1 to 276 of 276 are 94.56 pct identical to residues 1 to 276 from Escherichia coli K-12 : B3554"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAP19149.1"
 /db_xref="GI:30043429"
 /translation="MILFGLRRKGIILQACGLSLRQTRVRCRCLFLGERSKMATG
 KSCSRWAPLAALMVLSLSCFCFEGDQKAFIDFLQNTVMRSGERLPTLTADQKKG
 FGPVSYAILGYSOOVQAMDSGLPVDVSNVIRVPODYVTSQGPLRMNGSLGV
 LAOLQNAKLQADAAHSAKQSDLLKPVFQAKTKVVTTPADALQPLIPAAQTFTQOL
 VMVDYIAQQQTQVSFVANGIQFTSQASSEYNKLIAPLPAHQHQAQNTATATQ
 complement(2215. .3189)
 /gene="yiaB"
 /locus_tag="S4182"
 complement(2215. .3189)
 /gene="yiaB"
 /locus_tag="S4182"
 /function="putative enzyme; Not classified"
 /note="residues 1 to 324 of 324 are 94.13 pct identical to residues 5 to 328 of 328 from Escherichia coli K-12 : B3553"
 /codon_start=1
 /transl_table=11
 /product="putative dehydrogenase"
 /protein_id="AAP19150.1"
 /db_xref="GI:30043430"
 /translation="MKPSVILYKALPDLLQRLQAHFTVHQAIVANLSPQTVVQNAIFA
 EAGLGSNENVDAALEKMPKLTATSTISVGNDFVDALTAARKILIMHTPTVLTET
 VADTLMALVLTARVVAERKAEKAGTASIGDWTGTDVHHKTLGIVGMRIGMAL
 AQRVHGFNPMILYARHHEKEAERNARYCDLDTLQSDFDVCLILPLTDETHLIF
 GAQFAMKSSAIFINAGRGVVDENALIALQKGEIHAAGLDVFEQELSDVSPILLS
 MANVAVPHIGSATHTRYGNACAVDNLIDALQCKVEKNCVNFHAD"
 complement(3293. .3919)
 /gene="yiaD"
 /locus_tag="S4183"
 complement(3293. .3919)
 /gene="yiaD"
 /locus_tag="S4183"
 /function="putative membrane; Not classified"
 /note="residues 1 to 208 of 208 are 88.94 pct identical to residues 12 to 219 of 219 from Escherichia coli K-12 : B3552"
 /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein"
 /protein_id="AAP19151.1"
 /db_xref="GI:30043431"
 /translation="MSGALAVSGCTTNPYTGEEAGKSAIGAGLSIVGAGIGALSSS
 KKDRGKALIGAAAGLGGVGYMDVQAKLDRKMGTVSVTRSGDNILINPNP
 VTPSSGAPLKPAGANTLTGVAVLVKYPKTVANVIGVYDSTGCHDLNRLSQORADS
 VASALITQGVDSRITQTGLGPANPIASNSTAECKAQRNRVEITLSPL"
 4160. .6439
 /gene="bisc"
 /locus_tag="S4184"
 4160. .6439
 /gene="bisc"
 /locus_tag="S4184"
 /function="enzyme; Biosynthesis of cofactors, carriers: Biotin"
 /note="residues 21 to 759 of 759 are 98.78 pct identical

to residues 1 to 739 of 739 from Escherichia coli K-12 :
 B3551
 /codon_start=1
 /transl_table=11
 /product="biotin sulfoxide reductase"
 /protein_id="AAP19152.1"
 /db_xref="GI:30043432"
 /translation="MLVETGCTVFSSRGALATMENSLSQAVRDQVHSTNRVRFPWV
 RKGFASFPQGRSGDFVVRWDEALDLHQHQRIRIAYGPASIFAGSYGWRSN
 GVLKASTLLQRYMALAGYTHGLDYSTGAQAQIMFYVVGSGEYVQQQTSWPLVLEH
 SDVVVLSANPLNTKLTAMNASDQGLSYFSALRDSGKLLICIDPMSESTVDVFDGRM
 EWVAPHGCTDVALMGLTAHTLVENGWDEAFIACTTGAVAFASYLLGSDGLAKNAE
 WAASICVGAKIRELAAIENQNTLMAGCMQORQPGQKHWMIIVTLAALCOIGT
 PGGGFSSTFPAANGNPTRRASVLSWQSLPGGCDVADKIPVIRVERALENPGGAVQ
 HGNWRHFPDRIFWAGAGNFTTHQDNTLRIRAWQKPELVVISECTWAAAHADIV
 LATTSTFQNLWTGYSNQHLVPMQVPPYEARNDFDFAELSERWEKGYARF
 TEGASQQLWFTFYVARQSGASQVLPFAEFQANQIETEPNDSERFTRPADP
 CDRPLAPKATSGKIFISORIDYGVPCDCHPMLEDEWOGNAEPQLOVLISAV
 PARHLSDQNSYSSIRELYAVANRPPVTHIPDDAARGITEGDMVYVWNSQOILAGAV
 ISEGIKGVICIHGAMPDLDLTADGICKNGAVNVLEDFIPSSRLNGCAGNTYLAWL
 EKYNGPHLTATAPFPASS"
 complement (6408) .6848)
 /gene="viac"
 /locus tag="S4185"
 complement (6408) .6848)
 /gene="viac"
 /locus tag="S4185"
 /note="residues 1 to 146 of 146 are 97.94 pct identical to
 residues 1 to 146 of 146 from Escherichia coli K-12 :
 B3550"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAP19153.1"
 /db_xref="GI:30043433"
 /translation="MIRARQSELPAILLEWSTTWGHPFKKANYWRDCIPIURDAY
 EVYQKQPAIDFYRAQGFHIVDCAWQDETQLTPTWNSPVPVQTL"
 complement (6845) .7408)
 /gene="tag"
 /locus tag="S4186"
 complement (6845) .7408)
 /gene="tag"
 /locus tag="S4186"
 /function="enzyme; DNA - replication, repair,
 restriction/modification"
 /note="residues 1 to 187 of 187 are 97.32 pct identical to
 residues 1 to 187 of 187 from Escherichia coli K-12 :
 B3549"
 /codon_start=1
 /transl_table=11
 /product="3-methyl-adenine DNA glycosylase I"
 /protein_id="AAP19154.1"
 /db_xref="GI:30043434"
 /translation="MERCQWVSOGLPIYAHDNWGVPEPDSKKLFEMICFSGOQAGL
 SWITVKKRNVAYFHFQDPKVAAMQDEVERLVQDAGVIRHGKIQALIGNARAY
 IQMQRGEPFDPFVNSFVNHQPOVQATLSEIPTSTSSADLSALKKKGKFKVGT
 ICISFNAQGLVNDHVVGCCCYLGNKP"
 complement (7566) .8264
 /gene="yhjy"
 /locus tag="S4187"
 complement (7566) .8264
 /gene="yhjy"
 /locus tag="S4187"
 /function="putative enzyme; Not classified"
 /note="residues 1 to 232 of 232 are 99.13 pct identical to
 residues 1 to 232 of 232 from Escherichia coli K-12 :
 B3548"
 /codon_start=1
 /transl_table=11
 /product="glutathione oxidoreductase"
 /protein_id="AAP19155.1"
 /db_xref="GI:30043435"
 /translation="MTKHYDIAIGGSGGASINRAAMYGKQCALIEAKELGGTCVN
 VGCVPKVMWHAQIREAIHMYGDPDTINKFWETLIASRTAYIDSIHYSYEV
 LGKNNVDVIKGFARVDATLEVNGETITADHILIAIGKRPSPHDIPGVYSDSGF
 FALPALPVRVAVGAVYAVELAGVINGLAKTHLFPVKHAPLSPDPMISLIVEM
 NARGPOLHTNAIPKAVVKNADGSLTLEDRSETVDCLIAIGREPANDINLEAAG
 VKNTEKGYIVDKYQNTNVEGIYAVGDTGAVELTPVAVAGRPISERLFPNNKDEHL
 DYSNIPVVFSPKFTVGLTPEQAREQDDQVKYKSSFTAMYATVTHRQPCRMK
 XVCVGPREEKIVGHIGIPGMDENLQGFVAVALKMGATKDFDNTVALHPTAAEEFVTRK
 "
 complement (2300) .2653
 /gene="arsR"
 /locus tag="S4188"
 complement (2300) .2653
 /gene="arsR"

Db 66047 TTAAGTCATATATGTTTTGACTTA 66023

RESULT 8

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS


```

pieces.
ACCESSION BX545855
VERSION   BX545855.2 GI:33412513
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE    Danio rerio (zebrafish)
ORGANISM  Danio rerio
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
           Cypriniformes; Cyprinidae; Danio.
           1 (bases 1 to 207991)
McLaren,S.
REFERENCE 1
AUTHORS   McLaren,S.
TITLE     Direct Submission
JOURNAL   Submitted (31-JUL-2003) Wellcome Trust Sanger Institute, Hinxton,
           Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
           zfish-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
           On Aug 1, 2003 this sequence version replaced gi:32187728.
COMMENT   ----- Genome Center
           Center: Wellcome Trust Sanger Institute
           Center code: SC
           Web site: http://www.sanger.ac.uk
           Contact: zfish-help@sanger.ac.uk
           ----- Project Information
           Center project name: zc212N6
           ----- Summary Statistics
           Chemistry: Program: XGAP4; version 4.5
           Consensus quality: 205928 bases at least Q40
           Consensus quality: 206286 bases at least Q40
           Consensus quality: 206836 bases at least Q20
           Insert size: 207491; sum-of-contigs
           Insert size: 206582; 6.8% error; agarose-fp
           Quality coverage: 10.18x in Q20 bases; sum-of-contigs Quality
           coverage: 10.42x in Q20 bases; agarose-fp
           -----
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 6 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence.
           * as soon as it is available and the accession number will
           * be preserved.
           *
           * 1 33257: contig of 33257 bp in length
           * 33258 33357: gap of 100 bp
           * 33358 82317: contig of 48960 bp in length
           * 82318 82417: gap of 100 bp
           * 82418 129660: contig of 47243 bp in length
           * 129661 129760: gap of 100 bp
           * 129761 140386: contig of 10626 bp in length
           * 140387 140486: gap of 100 bp
           * 140487 185104: contig of 44618 bp in length
           * 185105 185204: gap of 100 bp
           * 185205 207991: contig of 22787 bp in length.
           Location/Qualifiers
           1..207991
            /organism="Danio rerio"
            /mol_type="genomic DNA"
            /db_xref="taxon:7955"
            /clone="CH211-212N6"
            /clone_lib="CHORI-211"
            1..33257
             /note="assembly fragment:00933
             fragment_chain:1
             clone_end:T7
             vector_side:left"
             33358..82317
              /note="assembly fragment:01153
              fragment_chain:1"
              82418..129660
               /note="assembly fragment:00434
               fragment_chain:2"
               129761..140386
                /note="assembly fragment:00494
                fragment_chain:2"

```

```

fragment_chain:2"
140487..185104
 /note="assembly fragment:02420
 fragment_chain:2"
185205..207991
 /note="assembly fragment:01206
 fragment_chain:2"
clone_end:SP6
vector_side:right"

ORIGIN
Query Match      87.3%; Score 21.8; DB 2; Length 207991;
Best Local Similarity 92.0%; Pred. No. 1e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TTAAGTCATATATGCTTTTTCACCTTA 25
    |||||
Db 190319 TTAATCATATATGCTTTTTCATTTA 190295

RESULT 10
LOCUS      AP002565                267888 bp    DNA    linear    BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 16/20.
ACCESSION  AP002565 BA000007
VERSION    AP002565.1 GI:13363693
KEYWORDS
SOURCE     Escherichia coli O157:H7
ORGANISM   Escherichia coli O157:H7
REFERENCE  1 (sites)
           Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
           Enterobacteriaceae; Escherichia.
           Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
           Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
           Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
           Sasakawa,C. and Shinagawa,H.
           Complete nucleotide sequence of the prophage VT2-Sakai carrying the
           verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
           derived from the Sakai outbreak
           Genes Genet. Syst. 74 (5), 227-239 (1999)
JOURNAL    MEDLINE
PUBMED    20198780
REFERENCE  2 (sites)
           10734605
AUTHORS    Ohnishi,M., Murata,T., Nakayama,K., Kubota,Y., Hattori,M.,
           Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
           Hayashi,T.
TITLE      Comparative analysis of the whole set of rRNA operons between an
           enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
           Escherichia coli K-12 strain MGI655
JOURNAL    Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
MEDLINE    20557356
PUBMED    11108008
REFERENCE  3 (sites)
           Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
           Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
           Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
           Shinagawa,H.
TITLE      Complete nucleotide sequence of the prophage VT1-Sakai carrying the
           Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
           O157:H7 strain derived from the Sakai outbreak
           Gene 258 (1-2), 127-139 (2000)
JOURNAL    MEDLINE
PUBMED    20564182
REFERENCE  4 (sites)
           1111050
AUTHORS    Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
           Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
           Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
           Ogasawara,N., Yasunaga,T., Kubota,Y., Shiba,T., Hattori,M. and
           Shinagawa,H.
TITLE      Complete genome sequence of enterohemorrhagic Escherichia coli
           O157:H7 and genomic comparison with a laboratory strain K-12
           DNA Res. 8 (1), 11-22 (2001)
JOURNAL    MEDLINE
PUBMED    21156231

```


/gene="EC4227"
/note="Similar to GPH ECOLI gi|1789787 percent identity 99
in 252 aa (Conserved in E.coli K-12)"

Query Match 87.2%; Score 21.8; DB 1; Length 267888;
Best Local Similarity 92.0%; Pred. No. 95;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTGGACTTA 25
Db 163563 TTAACCATATATGTTTGGACTTA 163587

RESULT 11
AL954643/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AL954643 71661 bp DNA linear ROD 27-JUN-2003
Mouse DNA sequence from clone RP24-142A19 on chromosome X, complete
sequence.
AL954643
AL954643.8 GI:32328853
HTG.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Wray, P.
1 (bases 1 to 71661)
Direct Submission
Submitted (26-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 27, 2003 this sequence version replaced gi:32131025.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP24-142A19 is
from a Male (C57BL/6J) mouse BAC Library VECTOR: P1A8BAC1.

FEATURES
source
1. .71661
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP24-142A19"
/clone_lib="RP24-142A19"

ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 71661;

Best Local Similarity 91.7%; Pred. No. 3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TTAAGTCATATATGTTTGGACTT 24
Db 68114 TTAAGTCATATATCTTTTGACTT 68091

RESULT 12
HS384F21/c
LOCUS
DEFINITION

HS384F21 96879 bp DNA linear PRI 05-JUN-2003
Human DNA sequence from clone RP3-384F21 on chromosome 1q24
Contains probable G protein-coupled receptor, EST, SFS, CA repeat,
complete sequence.
AL022171
AL022171.1 GI:3319684
HTG; repeat polymorphism.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Wilkinson, J.
1 (bases 1 to 96879)
Direct Submission
Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jul 15, 1998 this sequence version replaced gi:2980810.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

COMMENT

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGT/Chrl
RP3-384F21 is from the library RPCL-3 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm

VECTOR: PCWPAC2
IMPORTANT: This sequence is not the entire insert of clone
RP3-384F21. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true right end of clone RP3-384F21 is at 96879 in this
sequence. The true right end of clone RP4-809F4 is at 105 in this
sequence.

FEATURES
Location/Qualifiers
1. .96879
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="RZPD:RPCIP704F21384"
/db_xref="taxon:9606"

RESULT 13
AL954640
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AL954640 112484 bp DNA linear ROD 24-JAN-2003
Mouse DNA sequence from clone RP23-65C22 on chromosome X, complete
sequence.
AL954640
AL954640.6 GI:27899738
HTG.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 112484)
Heath, P.
Direct Submission
Submitted (24-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
On Jan 24, 2003 this sequence version replaced gi:27817477.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun map have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
from the RPI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBAC3.6.
Location/Qualifiers
1. .112484
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-65C22"
/clone_lib="RPCI-23"

FEATURES

Source
1. .112484
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-65C22"
/clone_lib="RPCI-23"

ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 112484;
Best Local Similarity 91.7%; Pred. No. 2.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TTAAGTCATATATGTTTTCACCTT 24
Db 68337 TTTAGTCATATATCTTTTTCACCTT 68360

RESULT 14

AC084813
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC084813 151950 bp DNA linear HTG 18-NOV-2000
Homo sapiens chromosome 8 clone RP11-21H16 map 8, WORKING DRAFT
SEQUENCE, 28 unordered pieces.
AC084813
AC084813.1 GI:11225414
HTG; HTGS PHASE1; HTGS_DRAFT.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 151950)
Biren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 8, clone RP11-21H16
Unpublished
2 (bases 1 to 151950)
Biren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bede, F., Boguski, L.,
Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
Choe, Y., Colling, M., Collins, S., Collymore, A., Cooke, P.,
DeRellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,
FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,
Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L.,
Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Lacroque, K.,
Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Liu, C., Liu, G.,
Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McKernan, K.,
McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,
Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,
O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riback, M., Riley, R.,
Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
Sougnaz, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (18-NOV-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRK
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L10793
Center clone name: 21_H16
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 137649 bases at least Q40
Consensus quality: 144628 bases at least Q30
Consensus quality: 147471 bases at least Q20
Insert size: 176000; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs
Quality coverage: 4.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 147: contig of 147 bp in length
* 148 247: gap of 100 bp
* 248 1339: contig of 1092 bp in length

```

* 1340 1439: gap of 100 bp
* 1440 2893: contig of 1454 bp in length
* 2894 2993: gap of 100 bp
* 2994 4018: contig of 1025 bp in length
* 4019 5251: contig of 1133 bp in length
* 5252 6793: contig of 1328 bp in length
* 6794 8351: contig of 1572 bp in length
* 8352 10117: contig of 1666 bp in length
* 10118 11508: contig of 1291 bp in length
* 11509 12787: contig of 1179 bp in length
* 12788 14292: contig of 1405 bp in length
* 14293 15575: contig of 1183 bp in length
* 15576 17805: contig of 2130 bp in length
* 17806 20308: contig of 2403 bp in length
* 20309 22679: contig of 2271 bp in length
* 22680 27233: contig of 4454 bp in length
* 27234 31001: contig of 3667 bp in length
* 31002 37219: contig of 6119 bp in length
* 37220 43086: contig of 5767 bp in length
* 43087 49376: contig of 6190 bp in length
* 49377 57190: contig of 7714 bp in length
* 57191 65900: contig of 8609 bp in length
* 65901 76412: contig of 10412 bp in length
* 76413 91478: contig of 14967 bp in length
* 91479 102755: contig of 11177 bp in length
* 102756 117027: contig of 14172 bp in length
* 117028 130830: contig of 13703 bp in length
* 130831 151950: contig of 21020 bp in length.
* 151951 151950: contig of 21020 bp in length.

```

FEATURES

```

source
1..151950
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-21H16"
/clone_lib="RPC11 Human Male BAC"
1..1147
/note="assembly_fragment"
vector side:left"
248..1339
/note="assembly_fragment"
1440..2893
/note="assembly_fragment"
2994..4018
/note="assembly_fragment"
4119..5251
/note="assembly_fragment"

```

misc_feature

```

1..1147
/note="assembly_fragment"
vector side:left"

```

misc_feature

```

248..1339
/note="assembly_fragment"

```

misc_feature

```

1440..2893
/note="assembly_fragment"

```

misc_feature

```

2994..4018
/note="assembly_fragment"

```

misc_feature

```

4119..5251
/note="assembly_fragment"

```

```

misc_feature 5352..6679
/note="assembly_fragment"
misc_feature 6780..8351
/note="assembly_fragment"
misc_feature 8452..10117
/note="assembly_fragment"
misc_feature 10218..11508
/note="assembly_fragment"
misc_feature 11609..12787
/note="assembly_fragment"
misc_feature 12888..14292
/note="assembly_fragment"
misc_feature 14393..15575
/note="assembly_fragment"
misc_feature 15676..17805
/note="assembly_fragment"
misc_feature 17906..20308
/note="assembly_fragment"
misc_feature 20409..22679
/note="assembly_fragment"
misc_feature 22780..27233
/note="assembly_fragment"
misc_feature 27334..31000
/note="assembly_fragment"
misc_feature 31101..37219
/note="assembly_fragment"
misc_feature 37320..43086
/note="assembly_fragment"
misc_feature 43187..49376
/note="assembly_fragment"
misc_feature 49477..57190
/note="assembly_fragment"
misc_feature 57291..65899
/note="assembly_fragment"
misc_feature 66000..76411
/note="assembly_fragment"
misc_feature 76512..91478
/note="assembly_fragment"
misc_feature 91579..102755
/note="assembly_fragment"
misc_feature 102856..117027
/note="assembly_fragment"
misc_feature 117128..130830
/note="assembly_fragment"
misc_feature 130931..151950
/note="assembly_fragment"

```

ORIGIN

Query March 83.2%; Score 20.8; DB 2; Length 151950;
 Best Local Similarity 91.7%; Pred. No. 2.5e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGCTTTTGACTT 24

Db 38807 TTTAGTCATATATCTTTTGACTT 38830

RESULT 15

AX508952/c 2000 bp DNA linear PAT 27-SEP-2002
 LOCUS AX508952
 DEFINITION Sequence 3647 from Patent WO0216655.
 ACCESSION AX508952
 VERSION AX508952.1 GI:23390189

KEYWORDS

ORGANISM
 Arabidopsis thaliana (thale cress)

Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE

1
 AUTHORS Harper, J.F., Kreps, J., Wang, X. and Zhu, T.
 TITLE Stress-regulated genes of plants, transgenic plants containing same, and methods of use

Search completed: May 26, 2004, 17:22:30
Job time : 353.756 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 15:20:44 ; Search time 347.756 Seconds
(without alignments)
3115.905 Million cell updates/sec

Title: US-10-676-299-10

Perfect score: 25

Sequence: 1 taagtcacaaacatatgacttaa 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rtd.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	25	100.0	3492	1	ECARSRBC	X80057 E.coli gene
C 2	25	100.0	10240	1	AE000426	AE000426 Escherich
C 3	25	100.0	11524	1	AE015361	AE015361 Shigella
C 4	25	100.0	179941	2	AC145934	AC145934 Gallus ga
C 5	25	100.0	225419	1	ECOUW76	U00039 E. coli chr
C 6	25	100.0	242495	2	AC146183	AC146183 Pan trogl
C 7	25	100.0	289816	1	AE016992	AE016992 Shigella
C 8	21.8	87.2	11071	1	AE005575	AE005575 Escherich
C 9	21.8	87.2	207991	2	BX545855	BX545855 Danio rer
C 10	21.8	87.2	267888	1	AP002565	AP002565 Escherich
C 11	20.8	83.2	71661	10	AL954643	AL954643 Mouse DNA
C 12	20.8	83.2	96879	9	HS384F21	AL022171 Human DNA
C 13	20.8	83.2	112484	10	AL954640	AL954640 Mouse DNA
C 14	20.8	83.2	151950	2	AC084813	AC084813 Homo sapi
C 15	20.2	80.8	2000	6	AX508952	AX508952 Sequence
C 16	20.2	80.8	5641	6	AX346299	AX346299 Sequence
C 17	20.2	80.8	89779	8	AB005234	AB005234 Arabidops
C 18	20.2	80.8	101371	8	AC002534	AC002534 Arabidops
C 19	20.2	80.8	146712	5	AL935136	AL935136 Zebrafish
C 20	20.2	80.8	182803	2	AC145899	AC145899 Pan trogl
C 21	20.2	80.8	186981	9	AC006459	AC006459 Homo sapi
C 22	20.2	80.8	216010	2	AC119517	AC119517 Rattus no
C 23	20.2	80.8	219476	9	AC068945	AC068945 Homo sapi
C 24	20.2	80.8	220414	2	BX546482	BX546482 Danio rer
C 25	20.2	80.8	240663	2	AC097541	AC097541 Rattus no
C 26	19.8	79.2	884	11	CNS058KTW	AL403338 T3 end of
C 27	19.8	79.2	57902	2	AC103790	AC103790 Homo sapi
C 28	19.8	79.2	68181	2	AC103792	AC103792 Homo sapi
C 29	19.8	79.2	94609	8	AB046439	AB046439 Arabidops
C 30	19.8	79.2	112615	9	AC079617	AC079617 Homo sapi
C 31	19.8	79.2	115626	9	AC103792	AC103792 Homo sapi
C 32	19.8	79.2	116461	8	AC051625	AC051625 Genomic S
C 33	19.8	79.2	121038	8	AP003342	AP003342 Oryza sat
C 34	19.8	79.2	121501	8	AC069557	AC069557 Genomic S
C 35	19.8	79.2	149098	8	AP003296	AP003296 Oryza sat
C 36	19.8	79.2	152813	2	AP005860	AP005860 Oryza sat
C 37	19.8	79.2	157295	2	AC079055	AC079055 Homo sapi
C 38	19.8	79.2	159302	2	AC098585	AC098585 Homo sapi
C 39	19.8	79.2	160295	10	AL845504	AL845504 Mouse DNA
C 40	19.8	79.2	160586	2	AC126319	AC126319 Mus muscu
C 41	19.8	79.2	176959	9	AC026402	AC026402 Homo sapi
C 42	19.8	79.2	182554	2	AC138307	AC138307 Mus muscu
C 43	19.8	79.2	183629	10	AC132596	AC132596 Mus muscu
C 44	19.8	79.2	189245	2	AC068090	AC068090 Homo sapi
C 45	19.8	79.2	195494	2	BX545912	BX545912 Danio rer

ALIGNMENTS

RESULT 1	ECARSRBC	3492 bp	DNA	linear	BCT 20-JUL-1995
ECARSRBC/c	E.coli genes arsr, arsb, arsc.				
LOCUS	X80057				
DEFINITION	X80057.1 GI:510824				
ACCESSION	arsB gene; arsc gene; arsenate reductase; arsenic-efflux pump;				
VERSION	arsenic-inducible repressor; arsr gene.				
KEYWORDS	Escherichia coli				
SOURCE	Escherichia coli				
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia				
REFERENCE	1				
AUTHORS	Diorio, C., Cai, J., Marmor, J., Shinder, R. and DuBow M.S.				
TITLE	An Escherichia coli chromosomal ars operon homolog is functional in				

Pred. No. is the number of results predicted by chance to have a

arsenic detoxification and is conserved in gram-negative bacteria
J. Bacteriol. 177 (8), 2050-2056 (1995)

95238276
7721697
2 (bases 1 to 3492)

Diorio, C.
Direct Submission
Submitted (06-JUL-1994) C. Diorio, McGill University, 3775
University Street, Montreal, Quebec, H3A 2B4, CANADA
Location/Qualifiers

1. 3492
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K12"

/db_xref="taxon:562"
528. 538
/rpt_type=INVERTED
544. 554
/rpt_type=INVERTED

655. 1061
/gene="arsR"
655. 660
/gene="arsR"
678. 683
/gene="arsR"

697. 701
/gene="arsR"
708. 1061
/gene="arsR"

/codon_start=1
/transl_table=11
/product="arsenic-inducible repressor"

/protein_id="CAA56361.1"
/db_xref="GI:510825"
/db_xref="GOA:P37309"
/db_xref="SWISS-PROT:P37309"

/translation="MSFLPIOLFKEIDETRLGVLISLSELGELCVCDLCTALDQSQ
PRISHALLRESGLLRLKQGWYHRLSPHAWAKIIDEAWRCOEKQVQALVRN
LARQNCSDSKNICS"

1071. 2404
/gene="arsB"
1071. 1076
/gene="arsB"

1072. 1084
/gene="arsB"
/rpt_type=INVERTED
1087. 1099
/rpt_type=INVERTED

1088. 1093
/gene="arsB"
1106. 1110
/gene="arsB"

1115. 2404
/gene="arsB"
/codon_start=1
/transl_table=11

/product="membrane-located arsenite efflux pump"
/protein_id="CAA56362.1"
/db_xref="GI:516211"

/db_xref="GOA:P37310"
/db_xref="SWISS-PROT:P37310"
/translation="MLLAGAIFVLITVLVWQKGLIGWSATLGAVALVGVVHPG
DIPVNVVNATAPVAVIITISLIDSGFFEMALHVSRWNGRGLLFTWIVLLG
NAYALFANDGALALITPVIAMLLALGFKTTLAFVWAGFIADTSLPLVSNLV
NIVASDFGLGFEYASVWVDIDAVIATLYMLHYFEKIDIPQNDWALLKSPARAI
KDPATKTKGWLVLILLVGFVLEPLGIPVSAIANGALLILFVAKRGHAIWTKVIR
GAPQVITFSLGLVLYVYGLRNLGTEYLSUGNLVADNGLWALITGTLFAPLSSI
MNMPTVLVGLSIDGSTASGVIKEMVYANVIGDLPKTIPIGSLATILMLHLVLSQ
KNITISGYYFRTGIITWLPVLVFLVTLAALALSLFL"

2417. 2842
/gene="arsC"
2417. 2842
/gene="arsC"

/codon_start=1
gene
CDS

ORIGIN

Query Match 100.0%; Score 25; DB 1; Length 3492;
Best Local Similarity 100.0%; Pred. NO. 14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 TAAGTCAAAAACATATATGACTTAA 25
|||||
Db 662 TAAGTCAAAAACATATATGACTTAA 638

repeat_region
repeat_region
rpt_type=INVERTED
rpt_type=INVERTED

AE000426
Escherichia coli K12 MG1655 section 316 of 400 of the complete
genome.

AE000426 U00096
AE000426.1 GI:1789910

Escherichia coli K12
Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

1 (bases 1 to 10240)
Blattner, F.R.; Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J.J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.

The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)
97426617
9278503

2 (bases 1 to 10240)
Blattner, F.R.
Direct Submission

Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459

3 (bases 1 to 10240)
Blattner, F.R.
Direct Submission

Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459

4 (bases 1 to 10240)
Direct Submission
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using GeneMark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 (e-mail: mark@marlab.gatech.edu). Open reading frames that
have been correlated with genetic loci are being annotated with CG

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

```

source
1..10240
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
<1..31
/promoter
/notes="factor Sigma70; predicted +1 start at 3638369"
161..1630
/gene="yhiP"
/notes="synonym: b3496"
161..1630
/gene="yhiP"
/function="putative transport; Not classified"
/notes="0489; 99 pct identical; amino acid sequence and
equal length to YHIP_ECOLI SW: P36837"
/codon_start=1
/transl_table=11
/product="putative transport protein"
/protein_id="AAC76521.1"
/db_xref="GI:1789911"
/transl_table=11
/translation="WNTTTPMCMLOQPPFPMIFVFLWREPGYVGVGLAVFVVKQ
LQPSQEQAVTFGAFAALYGLISGVGDHLLGKTIYVGLVLAIGTFMTGMSL
LKPDLIFALGAVNGLFRKANPASLLSKCPKDPRLDGAFTLFTMSINGSLAL
SLAPVADRFGYSVTNCGAGLIALLVYIACRGWVDIGSEDPFRMSFKLVL
LGSVVIYFVCAAMHNEVANLVILVSIYVITIIFFRAFLDKTRGNMFVAVLML
EAVFVILYAOPTSLNPFANINNVHEILGFSINPVFOALNPFVWLASPLIAGIT
HGNKGKDISPMKPTLGMFCSLGFLTAARAGMFAAAGLTSPAFIVLVYLFQSLG
ELPISNLGLAMIALVPOHLGFIKGMFPLTQAAFLGLGYVATTTAVPDNITDPLET
LPVITNFKIGLVILGVAVMVLLVFWPLKRMIAIPESH"
complement(1679..2431)
/gene="yhiQ"
/notes="synonym: b3497"
complement(1679..2431)
/function="orf; Unknown"
/notes="t250; 100 pct identical to 218 amino acids of
YHIQ_ECOLI SW: P37633 but has 32 additional N-ter
residues; gtg start, alternate starts possible"
/codon_start=1
/transl_table=11
/product="orf; hypothetical protein"
/protein_id="AAC76522.1"
/db_xref="GI:1789912"
/transl_table=11
/translation="MKICIDETGTGDGALSVLAARWGLEHEDNMLVLTPEHLEL
RKRDPEKLGIFVDFVGGMAHRRKFGGGRGEAVAKAVGIGKDYLPDWDVATAGRD
AFVLASVGRVRLERNPVVAALDDGLARGYADAEIGWLOERLQLIHASSLTALTD
ITPRQVVYLDPMFPHKQKSAVVKEMRVFQSLVGFPLDADGLLEPARLLATKRVVVK
RDPYAPPLANVATPNAVVTGHRFDIYAGTPV"
complement(2439..4481)
/gene="prlC"
/notes="synonym: b3498"
complement(2439..4481)
/gene="prlC"

```

```

/EC_number="3.4.24.70"
/function="enzyme; Degradation of proteins, peptides,
glyco"
/notes="f680; 100 pct identical to OPDA_ECOLI SW: P27298;
CG Site No. 18031"
/codon_start=1
/transl_table=11
/product="oligopeptidase A"
/protein_id="AAC76523.1"
/db_xref="GI:1789913"
/transl_table=11
/translation="WNTPLTTPPPLPPSKLPBHVPAVTKALMDCRENVERVVAQO
APYTWENLCQPLASVDVLAGRIFSPVSHLNSVKNPSPELRAYEOTLPILLSYSTWVGQ
HEGLYKAVRDLRDGDHYATLNTAOKKAVDNALRDFELSGIGLPEKQQRVGEIATRLS
ELNQYNNVLDATWGTWTKLYTDRALAGHPESALAAKAAKAEKELGYLLTLDIPS
LYPWTYCDNALREEMTRASTASDQGFNAGKWDNSKVMEHIALHRELALQLLQFGE
NYAPFKSLATKWAENPQVLDLTDLAKARVQGEKELAQRAFARAEKRGVDELQPDWI
FFEYKQKHLYSISDEQLRYPFNKAVNGLFVVGVIYGITAKRKDVGWHPDVR
PFYLEDNNELRSGFYLDLYARENKGAMDDCVGWRKADGSLQKQFVAYLTGCPNR
CMEPRALAFISCHYETGEPLKELLDKMLAKNYQALFILRQLEFGFLDFELHAEFR
PDQAKILLETALIKLVAVVPSWGRFPFAPSHIFAGGYAGGYAVYVLMADVLAADA
FSRFEKGFNRTGQSFLDNILSGGSEEPMDLTKFRPREPQDAMLEHYGKG"
complement(2457..2488)
/gene="prlC"
/notes="factor Sigma70; predicted +1 start at 3640781"
161..1630
/gene="prlC"
/notes="synonym: b3499"
161..1630
/gene="prlC"
/function="orf; Unknown"
/notes="0280; 100 pct identical; amino acid sequence and
equal length to YHIR_ECOLI SW: P37634"
/codon_start=1
/transl_table=11
/product="orf; hypothetical protein"
/protein_id="AAC76524.1"
/db_xref="GI:1789914"
/transl_table=11
/translation="WLSYRHSFHAGNEADVLKHTVQSLIIESLKEKDPFLYLDTHAG
AGRYQLASEHAEARTGEYLEGIARIWQDDLPABLEAYINVKHENSQLSYVPGSPL
IARLLRQDSLQUTELHPSDYLIRSEFQKDSARVAKADGFQQLKAKLPVSRRL
ILIDPPYEMKTDYQAVVSGTAEYKRFATYALWYPVLRQQIKRMHIDLEATGIRK
ILQIELAVLPDSRRRGWTASGMIVINPPWKLQEQMNVLPWLHSLKLVPAGTGATVSW
IVPS"
5291..5303
/gene="yhiR"
/notes="Central position to predicted promoter: -157.5"
/bound_moiety="Ada predicted site"
5421..5449
/gene="yhiR"
/notes="factor Sigma70; predicted +1 start at 3643787"
5598..6950
/gene="gor"
/notes="synonym: b3500"
5598..6950
/gene="gor"
/EC_numbers="1.6.4.2"
/function="enzyme; Biosynthesis of cofactors, carriers:
Thioredoxin, glutaredoxin, glutathione"
/notes="0450; 100 pct identical to GSHR_ECOLI SW: P06715;
CG Site No. 665"
/codon_start=1
/transl_table=11

```

66


```

* 16380 17948: contig of 1569 bp in length
* 17949 18048: gap of unknown length
* 18049 19612: contig of 1564 bp in length
* 19613 19712: gap of unknown length
* 19713 21723: contig of 2011 bp in length
* 21724 21823: gap of unknown length
* 21824 25266: contig of 3443 bp in length
* 25267 25366: gap of unknown length
* 25367 27105: contig of 1739 bp in length
* 27106 27205: gap of unknown length
* 27206 28336: contig of 1331 bp in length
* 28337 30905: gap of unknown length
* 30906 31005: contig of 2289 bp in length
* 31006 33224: contig of 2219 bp in length
* 33225 37384: gap of unknown length
* 37385 37484: contig of 4060 bp in length
* 37485 40797: gap of unknown length
* 40798 40897: contig of 3313 bp in length
* 40898 43117: gap of unknown length
* 43118 46243: contig of 2220 bp in length
* 46244 46343: gap of unknown length
* 46344 49765: contig of 3422 bp in length
* 49766 49865: gap of unknown length
* 49866 56038: contig of 6173 bp in length
* 56039 56138: gap of unknown length
* 56139 60188: contig of 4050 bp in length
* 60189 60288: gap of unknown length
* 60289 67106: contig of 6718 bp in length
* 67107 68878: gap of unknown length
* 68879 68978: gap of unknown length
* 68979 72905: contig of 3927 bp in length
* 72906 77513: gap of unknown length
* 77514 81177: contig of 3564 bp in length
* 81178 86122: gap of unknown length
* 86123 86712: gap of unknown length
* 86713 91513: contig of 4801 bp in length
* 91514 97537: contig of 5924 bp in length
* 97538 97637: gap of unknown length
* 97638 103312: contig of 5675 bp in length
* 103313 103412: gap of unknown length
* 103413 111607: contig of 8195 bp in length
* 111608 123016: contig of 11309 bp in length
* 123017 123116: gap of unknown length
* 123117 131760: contig of 8644 bp in length
* 131761 133264: gap of unknown length
* 133265 133664: contig of 1404 bp in length
* 133665 140148: gap of unknown length
* 140149 140248: gap of unknown length
* 140249 152045: contig of 11797 bp in length
* 152046 152145: gap of unknown length
* 152146 165190: contig of 13045 bp in length
* 165191 165290: gap of unknown length
* 165291 179941: contig of 14651 bp in length.
FEATURES
    location/Qualifiers
    1..179941
        /organism="Gallus gallus"
        /mol_type="genomic DNA"
        /db_xref="taxon:9031"
        /chromosome="DNK"
        /clone="CH261-4211"
    misc_feature
    1..1176
        /note="assembly_name:Contig10"
    misc_feature
    1277..2558

```

```

misc_feature
2659..3848
    /note="assembly_name:Contig17"
misc_feature
3949..5266
    /note="assembly_name:Contig25"
misc_feature
5367..6401
    /note="assembly_name:Contig30"
misc_feature
6502..7599
    /note="assembly_name:Contig31"
misc_feature
7700..8831
    /note="assembly_name:Contig32"
misc_feature
8932..10481
    /note="assembly_name:Contig36"
misc_feature
10582..12037
    /note="assembly_name:Contig39"
misc_feature
12138..13596
    /note="assembly_name:Contig40"
misc_feature
13697..14911
    /note="assembly_name:Contig41"
misc_feature
15012..16279
    /note="assembly_name:Contig42"
misc_feature
16380..17948
    /note="assembly_name:Contig43"
misc_feature
18049..19612
    /note="assembly_name:Contig44"
misc_feature
19713..21723
    /note="assembly_name:Contig45"
misc_feature
21824..25266
    /note="assembly_name:Contig47"
misc_feature
25367..27105
    /note="assembly_name:Contig48"
misc_feature
27206..28536
    /note="assembly_name:Contig49"
misc_feature
28637..30905
    /note="assembly_name:Contig50"
misc_feature
31006..33224
    /note="assembly_name:Contig51"
misc_feature
33325..37384
    /note="assembly_name:Contig52"
misc_feature
37485..40797
    /note="assembly_name:Contig53"
misc_feature
40898..43117
    /note="assembly_name:Contig54"
misc_feature
43218..46243
    /note="assembly_name:Contig55"
misc_feature
46344..49765
    /note="assembly_name:Contig56"
misc_feature
49866..56038
    /note="assembly_name:Contig57"
misc_feature
56139..60188
    /note="assembly_name:Contig58"
misc_feature
60289..67006
    /note="assembly_name:Contig59"
misc_feature
67107..68878
    /note="assembly_name:Contig6"
misc_feature
68979..72905
    /note="assembly_name:Contig60"
misc_feature
73006..77513
    /note="assembly_name:Contig61"
misc_feature
77614..81177
    /note="assembly_name:Contig62"
misc_feature
81278..86612
    /note="assembly_name:Contig63"
misc_feature
86713..91513
    /note="assembly_name:Contig64"
misc_feature
91614..97537

```

```

Query Match          100.0%; Score 25; DB 2; Length 179941;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Cyt 1 TAAGTCAAAACATATATGACTTAA 25
|||||

```

```
Db      132260 TAAGTCAAAAACATATATGACTTAA 132284

RESULT 5
ECOUW76/c
LOCUS   225419 bp      DNA      linear      BCT 07-NOV-1996
DEFINITION   E. coli chromosomal region from 76.0 to 81.5 minutes.
ACCESSION   U00039
VERSION     U00039.1  GI:466582
KEYWORDS
SOURCE
ORGANISM
Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1 (bases 1 to 225419)
Sofia,H.J., Burland,V., Daniels,D.L., Plunkett,G. III and
Biatner,F.R.
TITLE
Analysis of the Escherichia coli genome. V. DNA sequence of the
region from 76.0 to 81.5 minutes
Nucleic Acids Res. 22 (13), 2576-2586 (1994)
94316500
PUBMED
8041620
2 (bases 1 to 225419)
Plunkett,G. III.
Direct Subalssion
Submitted (25-MAR-1994) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
COMMENT
This sequence was determined as part of the E. coli Genome Project
(Frederick R. Blattner, director) at the University of
Wisconsin-Madison. Supported by award HG00301 from the NIH Human
Genome Project. The entire sequence was independently determined
from E. coli MG1655; overlaps and conflicts with other sequence
determinations are annotated. The start of this entry overlaps the
start of the entry ECOMW82 (L10328) by 547 bp.

FEATURES
             Location/Qualifiers
     source          1..225419
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K-12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:562"
     notes="This sequence comprises the following lambda
     clones: EC19-98, EC27-273, EC17-131, EC30MM5, EC30K609-3,
     EC27-334, EC30K606A-14, EC30K606A-12, EC30-154, EC27-385,
     EC27-1129, EC30MM9, EC27-239, EC30MM3, EC27-SF3955,
     EC18-200, EC27-1135, EC30R3900-8B, EC30R3900-5B, EC19MM6,
     EC30K573-3; M13mp19 or Janus vectors were used for
     subcloning"
     complement (<1..1221)
     /note="corresponds to M28722; ECOGGI(928..2148)"
     gene            complement (<1..1085)
                     /gene="ggt"
     CDS              complement (<1..1085)
                     /gene="ggt"
                     /EC_number="2.3.2.2"
                     /note="CG Site No. 18280; cold-regulated"
                     /codon_start=1
                     /transl_table=11
                     /product="gamma-glutamyltranspeptidase"
                     /protein_id="AAB18422.1"
                     /db_xref="GI:466583"
                     /translation="MIKPTFLRRVAIAALLSGSPSAAAPAPPVSYGVVERDVFHPV
                     RAKQWASVDATATQGVVDILKEGGNAVDAAVGYALAVTHPOAGNLGGGGFMILIR
                     SKNGTATIDFREMAPAKATRDMLDQGNPDSSKSLTSHLASGTGTVAGFSALDK
                     YGTPLNKVQPAKLARDGFIYNDALDDKTYGSEVLPHNHNSKAIIPWKEGEPLKK
                     GDTLVQANLAKSLKEMAEANGDEFFYGTIAEQAKQKNGGLITKEDLAAYKAVRT
                     PISGDYRGYGVYGNPPPSGGGIIHVQTLINLENFMKTYGFGSADAMQIMAEAKYAY
                     ADSRELYGDPDFVKVFWQALTNKAYAKS"
                     <60..15171
     misc_feature     /note="corresponds to lambda clone EC19-98"
     promoter         1067..1094

/note="39%"
complement(1085..1112)
/note="43%"
complement(1146..1173)
/note="45%"
complement(1181..2402)
/note="corresponds to X56908; ECUGPQQ(1..1222)"
1205..1645
/note="alternate gene name yhhA"
/codon_start=1
/transl_table=11
/label_ORF_ol146a
/db_xref="GI:466584"
/translation="MKRLILITALLPFYVFAQPIINTLNNPNQPGYQIPSQORWOTQMQ
TQOIQKMLNQKLTQLOQOHLNQINNSORVLOSQPCRPAPQAQMLNTNGG
MLNSNPDSINQHMPPERENGMLNQSPFPDPLKTIQ"
complement(1565..2383)
/note="corresponds to X14437; ECUGPQ(1..819)"
complement(1632..2375)
/gene="ugpQ"
complement(1632..2375)
/gene="ugpQ"
/EC_number="3.1.4.2"
/codon_start=1
/transl_table=11
/product="glycerophosphocholine phosphodiesterase"
/protein_id="AAB18424.1"
/db_xref="GI:466585"
/translation="MSNWPYPRIVAHRGGKGLAPENTILASIDVGAKYGHKMFDAKL
SKDEIFLHDDNLERTSGVAGELANQDLRYDAGSWYSKMGKGRPLLSQVAE
RCRCHGMANIEIKFTTGTGTLTGMVALAARELWAGWTPLLSSFEIDALEAQAQAA
PELPGLLDSDWRDRELTAREGCVSILHNHKLNKARVMQLKAGLRILVYTNKP
QRAELLGWGVCICTDAIDVIGNFTAQ"
1693..1727
/note="terminator-like sequence"
2192..2323
/note="predicted bend of 73.99 degrees"
complement(2225..6941)
/note="corresponds to X13141; ECUGP(1..4717)"
/gene="ugpC"
complement(2372..3481)
/gene="ugpC"
complement(2372..3481)
/gene="ugpC"
/function="sn-Glycerol-3-phosphate transport system"
/note="CG site No. 17794"
/transl_table=11
/protein_id="AAB18425.1"
/db_xref="GI:912455"
/translation="MPSCAWMSIVRNKAGLKLOAVTKSWDGTQVIKPLTLDVADGE
FTVMYGPSCKSTLLRMVAGLRYTEGDIINDQRTVTEMEPKDRGIAMVFONYALYP
HMSVENMANGLKIRGKQKQIAERVKAAARILELDGLKRPRLSGCORVAMGR
AIVRDPVAFVLPDEPLSLNDAKLRVQMRLELQQLHRLKITSILYTHDQVEMTLAQRV
MYMNGGVAEQIGTPTVEYKPSALFVAFISGPNMNLITGRVNNSTHFLPDGGIHEL
LANGYQVAGRMTLTGISRPEHIALSSQAGVPMVMDLTLEILGADNLAHGRWGQKLV
VRLAHOERPTAGSTLWLAENQLHLFDGTCQRV"
complement(3444..4289)
/gene="ugpB"
complement(3444..4289)
/gene="ugpB"
/function="sn-Glycerol-3-phosphate transport system"
/note="CG Site No. 17791; membrane protein"
/transl_table=11
/protein_id="AAB18426.1"
/db_xref="GI:466587"
/translation="MIENRPMITISHTMLIIGIANTLPLPYVFAATLDKQAVYAA
PNTLPGTHLENIHNIWNGVGTNSAPRMNLNSFVAFSITLKIIVSMLSAFAL
WNFRPFLRFLFWMIPTILMEFEVRIPTFVEVIANLOMLDSITAGLTPLMASATAP
LRFQFMTLPDLBVAARIDGASPMRPFCDIVFPLSKTLAALFVITFYGNQYLWP
LILITVDLGTTVAGIKGMIAATGEGTTENNSVAVAMLLITLIPPVIVLWQRAFRGL
VDSEK"
```


Query Match 100.0%; Score 25; DB 2; Length 242495;

Best Local Similarity 100.0%; Pred. No. 6;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAGTCAAAAACATATATGACTTAA 25

DB 3353 TAAGTCAAAAACATATATGACTTAA 3329

RESULT 7

LOCUS AE016992 289816 bp DNA linear BCT 22-APR-2003
DEFINITION Shigella flexneri 2a str. 2457T section 15 of 16 of the complete genome.

ACCESSION AE016992 AE014073

VERSION AE016992.1 GI:30043426

SOURCE Shigella flexneri 2a str. 2457T

KEYWORDS Shigella flexneri 2a str. 2457T

ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Shigella.

REFERENCE 1 (bases 1 to 289816)

AUTHORS Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W.,

Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A.,

Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S.,

Schwartz, D.C. and Blattner, F.R.

TITLE Complete Genome Sequence and Comparative Genomics of Shigella

flexneri Serotype 2a Strain 2457T

JOURNAL Infect. Immun. 71 (5), 2775-2786 (2003)

PUBMED 12704152

REFERENCE 2 (bases 1 to 289816)

AUTHORS Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W.,

Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A.,

Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S.,

Schwartz, D.C. and Blattner, F.R.

TITLE Direct Submission

JOURNAL Submitted (13-JUN-2002) Genetics Laboratory, University of

Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA

FEATURES

source

1..289816

/organism="Shigella flexneri 2a str. 2457T"

/mol_type="genomic DNA"

/strain="2457T"

/serotype="2a"

/db_xref="taxon:198215"

complement(238..450)

/gene="cspA"

/locus_tag="S4179"

complement(238..450)

/gene="cspA"

/locus_tag="S4179"

/function="regulator; Adaptations, atypical conditions"

/note="residues 1 to 70 of 70 are 100.00 pct identical to

residues 1 to 70 of 70 from Escherichia coli K-12 : B3556"

/codon_start=1

/transl_table=11

/product="cold shock protein 7.4, transcriptional

activator of hns"

/protein_id="AAP19147.1"

/db_xref="GI:30043427"

/translation="MSGKMTGIVKWFNADKGFITPDGSKDVFVHFSATQNDGYKS

LDEGQKVSFTESGAKGPAAGNVTSL"

complement(896..1021)

/gene="yiaG"

/locus_tag="S4180"

complement(896..1021)

/gene="yiaG"

/locus_tag="S4180"

/note="residues 1 to 41 of 41 are 100.00 pct identical to

residues 1 to 41 of 96 from Escherichia coli K-12 : B3555"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAP19148.1"

/db_xref="GI:30043428"

/translation="MEYKDPHMLSSLEQIVFKDETKKTLTHTTSCIEIOL"

1335..2165

/gene="yiaP"

/locus_tag="S4181"

1335..2165

/gene="yiaP"

/locus_tag="S4181"

/note="residues 1 to 276 of 276 are 94.56 pct identical to

residues 1 to 276 of 276 from Escherichia coli K-12 :

B3554"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAP19149.1"

/db_xref="GI:30043429"

/translation="MILPGLRRKGILOACPGISLSRQTRVCRCLFLGERSKQWATG

KSSRWTFAPLAALIMVSLSCFDKEDQKRAFIDFQNTVMRSGERLFTLTADKKQ

FGFVSDYALLYGYSQQVQNMDSGLSPVDSVNARVPQDYVTSQGLREMGSLGV

LAQQLQAKLQADAHSAKQSDLEKVFQATKVTVPADALQPLIPAAQTFOQL

VWGDYLAQQGTQVSPVANGIQPPTSQAASEYNKLIAPLPAQHQAFNAQTAVTATQ

"

complement(2215..3189)

/gene="yiaB"

/locus_tag="S4182"

complement(2215..3189)

/gene="yiaB"

/locus_tag="S4182"

/function="putative enzyme; Not classified"

/note="residues 1 to 324 of 324 are 94.13 pct identical to

residues 5 to 328 of 328 from Escherichia coli K-12 :

B3553"

/codon_start=1

/transl_table=11

/product="putative dehydrogenase"

/protein_id="AAP19150.1"

/db_xref="GI:30043430"

/translation="MKPSVILYKALPDLLQRLQAHFTVHGVANLSPTQVVQNAIPA

EAGLIGNSNVDALEKMPFKRATSTISVGINDFVDALTARKILLMHTVLTET

VADTLMALVLTARVVEVAERVAGWETASIGDWTGDSVDVHHTKILIGVGRIGMAL

AQRVHFQNNPILYNARHKEAEERFNARYCDLDTLLQSDFVCLILPTDETHLLP

GAEQFAKKSATFINAGRGVVDENALIAALQKEIHAAGLDVFEQESVSDSPILLS

MANYVAPHIGSATHETRYGMAACAVNDLIDALQCKVEKNCWPNHVA"

complement(3293..3919)

/gene="yiaD"

/locus_tag="S4183"

complement(3293..3919)

/gene="yiaD"

/locus_tag="S4183"

/function="putative membrane; Not classified"

/note="residues 1 to 208 of 208 are 88.94 pct identical to

residues 12 to 219 of 219 from Escherichia coli K-12 :

B3552"

/codon_start=1

/transl_table=11

/product="putative outer membrane protein"

/protein_id="AAP19151.1"

/db_xref="GI:30043431"

/translation="MSGALVSGCTNPTTCEREAGKSAICAGISLVGAGIGALSSS

KRQKRALGAAAGALGGGVGYMDVQAKLRDKRGKGVSVTRSGDNHILNPNPN

VTDFSSAPLKPAGANTLTGVAMVLYKEPTAVNVIGTSTGGHMLNRLSQQRAD

VASALITQGVDSARITQGLGPANPIASNSTAEKGAQNRRVEITLSPL"

4160..6439

/gene="bisc"

/locus_tag="S4184"

4160..6439

/gene="bisc"

/locus_tag="S4184"

/function="enzyme; Biosynthesis of cofactors, carriers;

Biotin"

/note="residues 21 to 759 of 759 are 98.78 pct identical

to residues 1 to 739 of 739 from Escherichia coli K-12 :
B3551"
/codon_start=1
/transl_table=11
/product="biotin sulfoxide reductase"
/protein_id="AAP19152.1"
/db_xref="GI:30043432"
/translation="MLVETDGETVPSSRGALATGKNSLQSAVDQVHSNTRVRFPVMV
RKGFASLSPNGQSGDFVVRVWDEALDIHQOHKRIEAYGPASIFAGSYGWRNS
GYLHASTLLQRYWALAGYTGHLGDTGAAQIMPTVVGSGSEYVQOQTSPWLVEH
SDVVLSANPLNTKILANWASDEGLSPGALRDSKKLICIDPMRSETVDFPDGKM
EWAPMGTVALKGIARTLVENGSHDEAFIARCTTGAVAFASVLLGESDGIANKAE
WAEICGVGAAGKIRLAI FQNTTLMAGWMQORQOGEOKHMMVITLAALGOJGT
FOGGFGLSHFANGNPTRSAVLSSMOGSLPGGCDVADKIPVARIVALENPGGAYQ
LNMNRHFPDIRIPWAGANFTHQDTRNLIRAQKQDELVI SECFWTAAKHADI V
LPATTSFERNLITGDSYNSQHLVPMQVPPVPEARNDFVFAELSERWKGXYARF
TEGSQLQWLETFTYVNAQRGASQOVELPPFAEFQANQLIEMPENPDSSEFIRADF
CRDPLAKLTAGSGIEIFSORIADYDPCDGHMMLPEDEWQCNPEPELOVLASAH
PAHRLHSOLNYSLLRELAVANREPVTHPDDAARGITEGDMVYVNSRGOILAGAV
ISEGIKRQVICHGAWPDLDTADGICKNGVNVVITXDLPSRLGNGCAGNTTLAWL
EKNYFPELITAFEPASS"
complement(6408..6848)
/gene="yiac"
/locus_tag="S4185"
complement(6408..6848)
/gene="yiac"
/locus_tag="S4185"
/note="Residues 1 to 146 of 146 are 97.94 pct identical to
residues 1 to 146 of 146 from Escherichia coli K-12 :
B3550"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP19153.1"
/db_xref="GI:30043433"
/translation="MIRAQRSELPAILLEWLESTTWGHPFIKANYWRDCIPLVRDAY
LANQWWEEDGKULGVISMEGRFLAMPVAPKVRGIGKALMQVQORYPHLM
EYQKNPAIDFYRAQGHIVDCAMQDETQPTMTSPVQVTL"
complement(6845..7408)
/gene="tag"
/locus_tag="S4186"
complement(6845..7408)
/gene="tag"
/locus_tag="S4186"
/function="enzyme; DNA - replication, repair,
restriction/modification"
/note="Residues 1 to 187 of 187 are 97.32 pct identical to
residues 1 to 187 of 187 from Escherichia coli K-12 :
B3549"
/codon_start=1
/transl_table=11
/product="3-methyl-adenine DNA glycosylase I"
/protein_id="AAP19154.1"
/db_xref="GI:30043434"
/translation="MERCQWVSGPLVYAHNDNGVVPETDSKKLFEMICFEGQAGL
SWITVTKRNYRAYFHQDPVKVAAQOEDEVERLVODAGIHRGKIQAILGNARAY
LQMEQNGQFPDFVASFVNHGPOVTOATLSEIFSTSTASDALSALKKRGFRVGT
ICISPMQACGLVNDHVVGCCCLGNKP"
7566..8264
/gene="yijy"
/locus_tag="S4187"
7566..8264
/gene="yijy"
/locus_tag="S4187"
/function="putative enzyme; Not classified"
/note="Residues 1 to 232 of 232 are 99.13 pct identical to
residues 1 to 232 of 232 from Escherichia coli K-12 :
B3550"
/codon_start=1
/transl_table=11
/product="glutathione oxidoreductase"
/protein_id="AAC58632.1"
/db_xref="GI:12518197"
/translation="MTKGYDYIAIGSGSGGIAINRAAMYQKCALIEAKELGTCVN
VGCNPKVWMAAQIREAHMYGPDYGFDTTKNFMTLIASRTAYIDRHSYENY
LGVNVDVVKGFARFVADAKTLENGVGETITADHLLIATGGRPSHPDIPGVYSDSGF
FALPALPERVAVAGVAVELAGVINGLCAKTHLFVRKHAPLRSFPMISLVEYVM
NAEGPOLHTVNAI PKAVVYADGSLTLELGRSETVDCLLWAGRPANINLEAAG
VTNEKGYIWDKQNTYVGIYAGDNTGAVELTPVAAGRISERLFPNKPDEHL
DYNIPFVFSHPFPIGTGLTEFQAREQVDDQVKKYSFTAMTATVTHRQPCRNK
XVCVGPSEKTVIHGIPGDMELQGFAPVAKMGATKKDPDNTVAIHPTAABEFVTNR
"
2300..2653
/gene="arsR"
/note="synonym: Z4903"
2300..2653
/gene="arsR"

Query Match 100.0%; Score 25; DB 1; Length 289816;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAAGTCAAAACATATAGACTTAA 25
|||||

Db 66023 TAAGTCAAAACATATATAGACTTAA 66047
RESULT 8
AE005575/c
LOCUS
DEFINITION
Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 194
of 290.
ACCESSION
AE005575.1 GI:12518196
VERSION
AE005575.1
KEYWORDS
SOURCE
ORGANISM
Escherichia coli O157:H7 EDL933
Escherichia coli O157:H7 EDL933
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 (bases 1 to 11071)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
21074935
PUBMED
11206551
REFERENCE
2 (bases 1 to 11071)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
LOCATION/Qualifiers
1..11071
/organism="Escherichia coli O157:H7 EDL933"
/mol_type="genomic DNA"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/notes="enterohaemorrhagic"
67..1419
/gene="gor"
/notes="synonym: Z4900"
67..1419
/gene="gor"
/function="enzyme; Biosynthesis of cofactors, carriers:
Thioredoxin, glutaredoxin, glutathione"
/note="Residues 1 to 450 of 450 are 99.33 pct identical to
residues 1 to 450 of 450 from Escherichia coli K-12 Strain
WG1655: B3500"
/codon_start=1
/transl_table=11
/product="glutathione oxidoreductase"
/protein_id="AAC58632.1"
/db_xref="GI:12518197"
/translation="MTKGYDYIAIGSGSGGIAINRAAMYQKCALIEAKELGTCVN
VGCNPKVWMAAQIREAHMYGPDYGFDTTKNFMTLIASRTAYIDRHSYENY
LGVNVDVVKGFARFVADAKTLENGVGETITADHLLIATGGRPSHPDIPGVYSDSGF
FALPALPERVAVAGVAVELAGVINGLCAKTHLFVRKHAPLRSFPMISLVEYVM
NAEGPOLHTVNAI PKAVVYADGSLTLELGRSETVDCLLWAGRPANINLEAAG
VTNEKGYIWDKQNTYVGIYAGDNTGAVELTPVAAGRISERLFPNKPDEHL
DYNIPFVFSHPFPIGTGLTEFQAREQVDDQVKKYSFTAMTATVTHRQPCRNK
XVCVGPSEKTVIHGIPGDMELQGFAPVAKMGATKKDPDNTVAIHPTAABEFVTNR
"
2300..2653
/gene="arsR"
/note="synonym: Z4903"
2300..2653
/gene="arsR"

```

/function="regulator; Protection responses: Drug/analog
sensitivity"
/note="Residues 1 to 117 of 117 are 94.87 pct identical to
residues 1 to 117 of 117 from Escherichia coli K-12 Strain
MG1655: B3501"
/codon_start=1
/transl_table=11
/product="transcriptional repressor of chromosomal ars
operon"
/protein_id="AAG58633.1"
/db_xref="GI:12518198"
/translation="NSFLPIOLFPIKILADETRLGIVLLSELGELVCVDLCTALDQSQ
PKISHLALLRSGLLRKQKQWHYRLSPHPSMAKIIIEQAWRCQEKVQVIVRN
LARQNSVDSKNTCS"
2707..3996
/gene="arsB"
/note="synonym: Z4904"
2707..3996
/gene="arsB"
/function="transport; Protection responses: Drug/analog
sensitivity"
/note="Residues 1 to 429 of 429 are 100.00 pct identical
to residues 8 to 436 of 436 from Escherichia coli K-12
Strain MG1655: B3502"
/codon_start=1
/transl_table=11
/product="arsenical pump membrane protein"
/protein_id="AAG58634.1"
/db_xref="GI:12518199"
/translation="MLLAGIPLVTLVWPKGLGIGWSATLGAVLALVTGVVHPG
DIPVVMIVNATAPIAVIIISLLDSGFFEAALVSRWNGRGLLFTWVLLG
AAVALFANDGAALITPIVIAMLLAGFSGKTITAFVMAAGFIADTASLPVLSLV
NIVSADFGLGVREVASVMPVDIAAIVATVMLHLVFRKQIPQNYDMALLKSPAEAI
KDPATPKTVGVVGLLLGLVLEPLGIPVSAIAVAGALILPVAKRGHAINVGKVLK
GAPQIVIPSLGMYLVVGLRNAGLTXVLSGLVNLADNGLWAATLGTGELTAPLSSI
KMNPTVLGALSIDSGVIRKENVANVIGDLGPKTPIGSLATILMLHLVLSQ
KMTISWGYPFTGIMTLPVFLVLAALALRLSFTL"
4009..4434
/gene="arsC"
/note="synonym: Z4905"
4009..4434
/gene="arsC"
/function="enzyme; Protection responses: Drug/analog
sensitivity"
/note="Residues 1 to 141 of 141 are 98.58 pct identical to
residues 1 to 141 of 141 from Escherichia coli K-12 Strain
MG1655: B3503"
/codon_start=1
/transl_table=11
/product="arsenate reductase"
/protein_id="AAG58635.1"
/db_xref="GI:12518200"
/translation="MSNITVYHNPACGTSRNTLEMRNSGTPTTIHYLETPTPRDEL
VKLIADMGISVRLARKNVEPYEELGLADKFTDRLIDFMLQHPILINPIVWTTIG
TLCRPSVVLLELPDAQGATTKEDGXVDBAGKRLK"
4546..4653
/gene="Z4906"
4546..4653
/gene="Z4906"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG58636.1"
/db_xref="GI:12518201"
/translation="MLFVGLTGVVHAQAILLFTHYSUREIPTIASSMMYQ"
5062..6177
/gene="Z4907"
5062..6177
/gene="Z4907"
/function="orf; Unknown function"
/note="Residues 1 to 249 of 371 are 89.95 pct identical to
residues 1 to 249 of 260 from Escherichia coli K-12 Strain

```

```

MG1655: B3504"
/codon_start=1
/transl_table=11
/protein_id="AAG58637.1"
/db_xref="GI:12518202"
/translation="MSIDFTPGMINTYHGDIVNRTTDTDNVKTPTDTPMPWCNREEQQ
PINTPSGEGVNPQYDLAROHLPQIVACRNTKYTDADYSKVAQLVSLITNIESIS
STPLTROORILNOINNIYKNSASCRILIVANPKDKALITKISVEGIPVRFPSV
QTMESDNFIABQADLPNTKIDJOSLYQKMTLYIHSSEKNRMKVAFAGTNPIDFM
QTQNLSGVFLTSLRFYFDLLNINFTDANLGDITFISHEHETPKLYKDEQVLDKQIES
LFTSLINDNFLEAKAEIATIIIFLEARTINLSYDILKYQQEFQKCYAQVKSFT
TLSYNIQIOWAEMSYQFEVYFYTLLNPKKCHARLI"
6499..7098
/gene="slp"
/note="synonym: Z4908"
6499..7098
/gene="slp"
/function="membrane; Cell envelop: Outer membrane
constituents"
/note="Residues 1 to 199 of 199 are 99.49 pct identical to
residues 1 to 199 of 199 from Escherichia coli K-12 Strain
MG1655: B3506"
/codon_start=1
/transl_table=11
/product="outer membrane protein induced after carbon
starvation"
/protein_id="AAG58638.1"
/db_xref="GI:12518203"
/translation="MWLIYVDKDSNMNMTKGALILSLFLLAACSSIPQNIKNNQPD
IQKSFVAVHQPLGVQQOAFEGKGVINVTDLLEIAVLPLDSYAKPDIEANYQ
GRLLARSGFDLPYRNHRHFTIIGTQGEOPGINKVPYFVFNFLSVNMQIQVHLEAV
VNTIYNWDYGYGAPWEPFGMGAPYYTNAVSOVTPELWK"
7254..7784
/gene="yhiP"
/note="synonym: Z4909"
7254..7784
/gene="yhiP"
/function="orf; Unknown function"
/note="Residues 1 to 176 of 176 are 99.43 pct identical to
residues 1 to 176 of 176 from Escherichia coli K-12 Strain
MG1655: B3507"
/codon_start=1
/transl_table=11
/product="orf; hypothetical protein"
/protein_id="AAG58639.1"
/db_xref="GI:12518204"
/translation="MFLIITRDTPFTAMKNILSKGNVHIIQNEBEEIDVMLHQAFAVI
IDTLNMYVHSNFTQIERLKPVHVIIFSPENIKRCIGKVPVTFVPTITIIDFVALL
NGSYCSVPEANVLSRKQHVLSICANQMTTIDLEKLKISLKYFCHKNIMMILNI
KRINELVRHQHIDYLV"
7786..>11071
/note="O-island #140; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
/complement(7845..8873)
/gene="chus"
/note="synonym: Z4910"
/complement(7845..8873)
/gene="chus"
/function="putative transport; Transport of small

```

Query Match 87.2%; Score 21.8; DB 1; Length 11071;
 Best Local Similarity 92.0%; Pred. No. 1.8e-02;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAAGTCAAAACATATATGACTTAA 25
 |||||
 DB 2254 TAAGTCAAAACATATATGCTTAA 2230

RESULT 9
 BX545855 207991 bp DNA linear HTG 01-AUG-2003
 LOCUS BX545855
 DEFINITION Danio rerio clone CH211-212N6, WORKING DRAFT SEQUENCE, 6 unordered


```

PUBMED 11258796
REFERENCE 5 (bases 1 to 267888)
AUTHORS Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Rayashi,T.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan [E-mail:kengen-info.osaka-u.ac.jp,
URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
Fax:81-6-6879-2047]
COMMENT genome project.
FEATURES Location/Qualifiers
source 1..267888
/organism="Escherichia coli O157:H7"
/mol_type="genomic DNA"
/strain="O157:H7"
/sub_strain="RIMD 0503952"
/db_xref="taxon:83334"
151..318
/gene="ECs4220"
151..318
/gene="ECs4220"
151..318
/notes="identical to YHFP_ECOLI gi|1789770 (Conserved in
E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37643.1"
/db_xref="GI:13363694"
/translation="MNKFIKVALGVAVLATLACTGHIENRDKNCSDYLLHPAISIS
KIIGCGPTAQ"
563..1951
/gene="ECs4221"
563..1951
/gene="ECs4221"
/notes="similar to YHFM_ECOLI gi|2367216 percent identity
99 in 462 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative amino acid/amine transport protein"
/protein_id="BAB37644.1"
/db_xref="GI:13363695"
/translation="MTANSPLORVGKGIAMGSOBLQKLGFWALAIAGTVTGSG
IFVSEVAKAAGTPMLTULAFVIGLIVIPOMCVVAELSTAVPENGADYVYLKNAGS
RLAPLUGSWAFWANDAPSUSIMALAIIVNMLGFLTIDPLLGKFIAGGLIAPMLLHL
RSVEGGAFTLITIAKIIPFTIVIGIGIFWKAENPAPATTAIGATSPHALLAGI
SATSWYTGMASTCYMTGEIKNPKTPRALIGSCLLVLYTLALVLSGLMPPDKL
ANSETISDALTIWIPALGSTAGIFVAITAMIVILGSLSCWYQPLEYAMAKNLF
KCFGHVHPKNTPDVSIILQOALGIFIPVSDLTSLIGVFLVNCFNLTIFGSIWC
EKDDYKPLARTPAFLMTPLATASSLILVASTFVWAPIPGLICAVIATGUPAYAF
WAKRSQNLALS"
1951..2994
/gene="ECs4222"
1951..2994
/gene="ECs4222"
/notes="similar to YHFN_ECOLI gi|1789772 percent identity
98 in 347 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative transport protein"
/protein_id="BAB37645.1"
/db_xref="GI:13363696"
/translation="MLSGENKMLDIDKSTVDFLVTENMQVEKVLSHDVLVTHIVE
EYKRDRLIVFVACGSLNAOATAKHLADRFSDLOVYASGWFCDNTPYRLDARCA
VIGSDYKTEEVKALBELGACGALTAAETKRAADSPITSAEFSIDYQADCTWEIHL
LLCVSVLEMLITLAPAEIGKIKNDLKLQPLNALCYLIVRTWEKQGLSLAQWPMI
YVAAGPLRLIGYKEGIVLMEFTWTHGCVESGEFPHGPIEIVFVWFLFLGNDZ
SRHTTERRAINFVKQRTDNIVIVDAISQGLHPWLPAPLFPVFWMLCYLYLSIYKDNH
PDERRYYGGLVEY"
3045..3872
/gene="ECs4223"
3045..3872
/gene="ECs4223"
/notes="N-terminal part (1-115 in 294 aa) is similar to
YHFO_ECOLI gi|1789773 percent identity 97 in 115 aa,
C-terminal part (162-294 in 294 aa) is similar to
YHFP_ECOLI gi|2367217 percent identity 99 in 133 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37646.1"
/db_xref="GI:13363697"
/translation="MKYGMFTCGHRLPIEHAFRDASLGYDGIEMGGRPHAFAPDL
KAGGIKQKALAQTYQMEFIITETETNGYPYMMGLGDRMRBSLDMLKLANDMAKEM
NAGYTLISAHAGYLTLPNNVIGRLAENLSICETAEINIGMDLILEPLTFVESVVCN
ANDVLEALVPSPLPMSVDICAPYQVAEPVMSYFDKLGDKLRHLHIVDSGADTH
YIPGEGKPLRELARDIIRDYEGYCTVELVTMYNPEPLRYARQALERFRALLPEDE"
3872..4204
/gene="ECs4224"
3872..4204
/gene="ECs4224"
/notes="similar to N-terminal part (1-110 in 261 aa) of
YHFO_ECOLI gi|1789775 percent identity 98 in 110 aa
(Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37647.1"
/db_xref="GI:13363698"
/translation="MKTLATIGDNCVDIYPQLNKAFFSGGNVAVNYCTRYGKPGCCI
TWVGDDYGTGLKQDLASMGVDISHVHTKHGVTAGTQVELHDNDRVFGDYTGVMADF
ALSEEDYA"
4691..5488
/gene="ECs4225"
4691..5488
/gene="ECs4225"
/notes="similar to YHFR_ECOLI gi|1789776 percent identity
98 in 265 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="BAB37648.1"
/db_xref="GI:13363699"
/translation="MLLAGKRPYNPCNSFFISGSIMSATDRYSHQLLYATVRQLLD
DIAQGVYAGQQIPTEBELCTQYNSRITIRKAIISDLVADGLVLRQKGTFFVQSKV
ENALVTSGFTDFGVSGKATKEKVIQERISAAPECKLINIPGNSEVHLCRVWYLD
KEPLIDSWSPLSRYPDELYVSGSTYQLQPERFTRVYSDKXTIDIFAATRPQA
KMKCELGEPLPRISKIAPFDQDKPKFVHSELFCRANRITLTIDNKEH"
5489..6623
/gene="ECs4226"
5489..6623
/gene="ECs4226"
/notes="similar to TRPS_ECOLI gi|1789786 percent identity
99 in 334 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="tryptophan tRNA synthetase"
/protein_id="BAB37649.1"
/db_xref="GI:13363700"
/translation="MTKPIVPSGAQPSGELTIGNYMGALRWQVNMQDHYHCYIVDQ
HAITVQDQKLRKATLTIALYLAGGIDPEKSTFVQSHVPEHAOLGHALNCTYVFG
ELSRMTPQDKARVARNAINAGLFDFVLMAADILLYQTNLPVGDQKQHLSESDI
AQRFNALYGDIFPKVPEPFIKSGARVMSLLELISAVTQGSIPLEKQFEGKYGHLKGE
KIKRVYTDDEPVPVYQVNGAGVSNLLDILSAVTQGSIPLEKQFEGKYGHLKGE
VADAVSGMLTELQERYHFRFENDEAFLOQVMDGAEKASASRSLKAVYEAIGFVAKP"
6624..7374
/complement (6616..7374)
7375..7374
/gene="ECs4227"
7375..7374
/complement (6616..7374)

```

/gene="ECs4227"
/note="similar to GPH_ECOLI gi11789787 percent identity 99
in 252 aa (Conserved in E.coli K-12)"

Query Match 87.2%; Score 21.8; DB 1; Length 267888;
Best Local Similarity 92.0%; Pred. No. 95;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAAGTCARAAACATATATGACTTAA 25
|||||
Db 163587 TAAGTCARAAACATATATGCTTAA 163563
|||||

RESULT 11
AL954643 71661 bp DNA linear ROD 27-JUN-2003
LOCUS Mouse DNA sequence from clone RP24-142A19 on chromosome X, complete
DEFINITION sequence.
ACCESSION AL954643
VERSION AL954643.8 GI:32328853
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 71661)
JOURNAL Wray.P.
COMMENT Direct Submission
Submitted (26-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 27, 2003 this sequence version replaced gi:32131025.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep RP24-142A19 is
from a Male (C57BL/6J) mouse BAC Library VECTOR: pTARBAC1.
Location/Qualifiers
1. 71661
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP24-142A19"
/clone_lib="RPC1-24"

FEATURES
source
1. 71661
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP24-142A19"
/clone_lib="RPC1-24"

Query Match 83.2%; Score 20.8; DB 10; Length 71661;

Best Local Similarity 91.7%; Pred. No. 3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 AAGTCARAAACATATATGACTTAA 25
|||||
Db 68091 AAGTCARAAACATATATGACTTAA 68114
|||||

RESULT 12
HS384F21 96879 bp DNA linear PRI 05-JUN-2003
LOCUS Human DNA sequence from clone RP3-384F21 on chromosome 1q24
DEFINITION contains probable G protein-coupled receptor, EST, STS, CA repeat,
complete sequence.
ACCESSION AL022171
VERSION AL022171.1 GI:3319684
KEYWORDS HTG; repeat polymorphism.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 96879)
JOURNAL Wilkinson.J.
COMMENT Direct Submission
Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jul 15, 1998 this sequence version replaced gi:2980810.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP3-384F21 is from the library RPC1-3 constructed by the group of
Pietter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCVPAC2

IMPORTANT: This sequence is not the entire insert of clone
RP3-384F21. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true right end of clone RP3-384F21 is at 96879 in this
sequence. The true right end of clone RP4-809F4 is at 105 in this
sequence.

FEATURES
Location/Qualifiers
1. 96879
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="R2PD:RPC1P704F21384"
/db_xref="taxon:9606"


```

RESULT 13
AL954640/c
LOCUS      112484 bp      DNA      linear      ROD 24-JAN-2003
DEFINITION Mouse DNA sequence from clone RP23-65C22 on chromosome X, complete
sequence.
ACCESSION  AL954640
VERSION     AL954640.6  GI:27899738
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus

REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE       Heath, P.
JOURNAL     Direct Submission
COMMENT     Submitted (24-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
           Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
           humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
           On Jan 24, 2003 this sequence version replaced GI:27817477.
           Sequence from the Mouse Genome Sequencing Consortium whole genome
           shotgun may have been used to confirm this sequence. Sequence data
           from the whole genome shotgun alone has only been used where it has
           a phred quality of at least 30.
           ----- Genome Center
           Center: Wellcome Trust Sanger Institute
           Center code: SC
           Web site: http://www.sanger.ac.uk
           Contact: humquery@sanger.ac.uk
           -----
           During sequence assembly data is compared from overlapping clones.
           Where differences are found these are annotated as variations
           together with a note of the overlapping clone name. Note that the
           variation annotation may not be found in the sequence submissions
           corresponding to the overlapping clone, as we submit sequences with
           only a small overlap as described above.
           This sequence was finished as follows unless otherwise noted: all
           regions were either double-stranded or sequenced with an alternate
           chemistry or covered by high quality data (i.e., phred quality >=
           30); an attempt was made to resolve all sequencing problems, such
           as compressions and repeats; all regions were covered by at least
           one plasmid subclone or more than one M13 subclone; and the
           assembly was confirmed by restriction digest, except on the rare
           occasion of the clone being a YAC.
           The following abbreviations are used to associate primary accession
           numbers given in the feature table with their source databases:
           Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
           on the WORMPEP database can be found at
           http://www.sanger.ac.uk/projects/C.elegans/wormpep
           from the RPI-23 Mouse BAC Library
           constructed by the group of Pieter de Jong.
           For further details see http://www.chori.org/bacpac/home.htm
           VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source           1..112484
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="genomic:10090"
                     /chromosome="X"
                     /clone="RP23-65C22"
                     /clone_lib="RPI-23"

ORIGIN
Query Match      83.2%; Score 20.8; DB 10; Length 112484;
Best Local Similarity 91.7%; Pred. No. 2.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      2  AAGTCAAAACATATATGACTTAA 25
          |||||
DB      68360  AAGTCAAAACATATATGACTTAA 68337
          |||||

RESULT 14
AC084813/c
LOCUS      151950 bp      DNA      linear      HTG 18-NOV-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-21H16 map 8, WORKING DRAFT
SEQUENCE, 28 unordered pieces.
ACCESSION  AC084813
VERSION     AC084813.1  GI:11225414
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens

REFERENCE   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS     1 (bases 1 to 151950)
TITLE       Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL     Unpublished
COMMENT     2 (bases 1 to 151950)
           Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
           Anderson, S., Barna, N., Bastien, V., Beda, F., Boguslavsky, L.,
           Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
           Deapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
           Dearth, K., Dewar, K., Diaz, J. S., Dodge, S., Ferreira, P.,
           FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,
           Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L.,
           Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Lacroque, K.,
           Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G.,
           Macdonald, P., Marquis, N., McCarthy, M., McGowan, P., McKernan, K.,
           McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
           Morrow, J., Murphy, I., Naylor, J., Norman, C. H., O'Connor, T.,
           O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K.,
           Pierre, N., Pisani, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,
           Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
           Sougnaz, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
           Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
           Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
           Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
           Zimmer, A. and Zody, M.
           Direct Submission
           Submitted (18-NOV-2000) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
           All repeats were identified using RepeatMasker:
           Smit, A. P. A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           ----- Genome Center
           Center: Whitehead Institute/ MIT Center for Genome Research
           Center code: WBIR
           Web site: http://www-seq.wi.mit.edu
           Contact: sequence_submissions@genome.wi.mit.edu
           ----- Project Information
           Center project name: L10793
           Center clone name: 21_H16
           ----- Summary Statistics
           Sequencing vector: Plasmid; n/a; 100% of reads
           Chemistry: Dye-terminator Big Dye; 100% of reads
           Assembly program: Phrap; version 0.960731
           Consensus quality: 137649 bases at least Q40
           Consensus quality: 144628 bases at least Q30
           Consensus quality: 147471 bases at least Q20
           Insert size: 176000; agarose-fp
           Quality coverage: 3.9 in Q20 bases; sum-of-contigs
           Quality coverage: 4.6 in Q20 bases; sum-of-contigs
           -----
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 28 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           * 1 147: contig of 147 bp in length
           * 148 247: gap of 100 bp
           * 248 1339: contig of 1092 bp in length

```

```

* 1340 1439: gap of 100 bp
* 1440 2893: contig of 1454 bp in length
* 2894 2993: gap of 100 bp
* 2994 4018: contig of 1025 bp in length
* 4019 5251: gap of 100 bp
* 5252 5351: contig of 1133 bp in length
* 5352 6679: gap of 100 bp
* 6680 6779: contig of 1328 bp in length
* 6780 8351: gap of 100 bp
* 8352 8451: contig of 1572 bp in length
* 8452 10117: gap of 100 bp
* 10118 10217: contig of 1666 bp in length
* 10219 11508: contig of 1291 bp in length
* 11509 12887: gap of 100 bp
* 12888 12897: contig of 1179 bp in length
* 12898 14292: gap of 100 bp
* 14293 14392: contig of 1405 bp in length
* 14393 15575: gap of 100 bp
* 15576 15675: contig of 1183 bp in length
* 15676 17805: gap of 100 bp
* 17806 17905: contig of 2130 bp in length
* 17906 20308: gap of 100 bp
* 20309 20409: contig of 2403 bp in length
* 20410 22679: gap of 100 bp
* 22680 22779: contig of 2271 bp in length
* 22780 27233: gap of 100 bp
* 27234 31000: contig of 4454 bp in length
* 31001 31100: gap of 100 bp
* 31101 37219: contig of 3667 bp in length
* 37220 37319: gap of 100 bp
* 37320 43086: contig of 6119 bp in length
* 43087 43186: gap of 100 bp
* 43187 49376: contig of 5767 bp in length
* 49377 49476: gap of 100 bp
* 49477 57190: contig of 6190 bp in length
* 57191 57290: gap of 100 bp
* 57291 58399: contig of 7714 bp in length
* 58400 65999: gap of 100 bp
* 66000 76411: contig of 14967 bp in length
* 76412 76511: gap of 100 bp
* 76512 91478: contig of 10412 bp in length
* 91479 91578: gap of 100 bp
* 91579 102755: contig of 14967 bp in length
* 102756 102855: gap of 100 bp
* 102856 117027: contig of 11177 bp in length
* 117028 117127: gap of 100 bp
* 117128 130830: contig of 14172 bp in length
* 130831 130930: gap of 100 bp
* 130931 151950: contig of 13703 bp in length
* 151951 151950: contig of 100 bp
* 151951 151950: contig of 21020 bp in length.

```

FEATURES

```

source
1. .151950
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-21H16"
/clone_lib="RPC1-11 Human Male BAC"
1. .147
/notes="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature
248..1339
/notes="assembly_fragment"
1440..2893
/notes="assembly_fragment"
2894..4018
/notes="assembly_fragment"
4019..5251
/notes="assembly_fragment"
5252..6679
/notes="assembly_fragment"
6680..8351
/notes="assembly_fragment"
8352..10117
/notes="assembly_fragment"
10118..10217
/notes="assembly_fragment"
10219..11508
/notes="assembly_fragment"
11509..12887
/notes="assembly_fragment"
12888..14292
/notes="assembly_fragment"
14293..15575
/notes="assembly_fragment"
15576..17805
/notes="assembly_fragment"
17806..20308
/notes="assembly_fragment"
20309..20409
/notes="assembly_fragment"
20410..22679
/notes="assembly_fragment"
22680..27233
/notes="assembly_fragment"
27234..31000
/notes="assembly_fragment"
31001..37219
/notes="assembly_fragment"
37220..43086
/notes="assembly_fragment"
43087..49376
/notes="assembly_fragment"
49377..57190
/notes="assembly_fragment"
57191..57290
/notes="assembly_fragment"
57291..58399
/notes="assembly_fragment"
58400..65999
/notes="assembly_fragment"
66000..76411
/notes="assembly_fragment"
76412..91478
/notes="assembly_fragment"
91479..91578
/notes="assembly_fragment"
91579..102755
/notes="assembly_fragment"
102756..102855
/notes="assembly_fragment"
102856..117027
/notes="assembly_fragment"
117028..130830
/notes="assembly_fragment"
130831..151950
/notes="assembly_fragment"

```

```

misc_feature
5352..6679
/notes="assembly_fragment"
6680..8351
/notes="assembly_fragment"
8352..10117
/notes="assembly_fragment"
10118..11508
/notes="assembly_fragment"
11509..12887
/notes="assembly_fragment"
12888..14292
/notes="assembly_fragment"
14293..15575
/notes="assembly_fragment"
15576..17805
/notes="assembly_fragment"
17806..20308
/notes="assembly_fragment"
20309..22679
/notes="assembly_fragment"
22680..27233
/notes="assembly_fragment"
27234..31000
/notes="assembly_fragment"
31001..37219
/notes="assembly_fragment"
37220..43086
/notes="assembly_fragment"
43087..49376
/notes="assembly_fragment"
49377..57190
/notes="assembly_fragment"
57191..57290
/notes="assembly_fragment"
57291..58399
/notes="assembly_fragment"
58400..65999
/notes="assembly_fragment"
66000..76411
/notes="assembly_fragment"
76412..91478
/notes="assembly_fragment"
91479..91578
/notes="assembly_fragment"
91579..102755
/notes="assembly_fragment"
102756..117027
/notes="assembly_fragment"
117028..130830
/notes="assembly_fragment"
130831..151950
/notes="assembly_fragment"

```

ORIGIN

```

Query Match      83.2%; Score 20.8; DB 2; Length 151950;
Best Local Similarity 91.7%; Pred. No. 2.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 AAGTCAAAACATATATGACTTAA 25
DB      38830 AAGTCAAAAGATATATGACTAAA 38807

```

RESULT 15

```

AX508952      2000 bp      DNA      linear      PAT 27-SEP-2002
LOCUS      AX508952
DEFINITION      Sequence 3647 from Patent WO0216655.
ACCESSION      AX508952
VERSION      AX508952.1
KEYWORDS      GI:23390189
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM      Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE      1
AUTHORS      Harper, J.F., Kreps, J., Wang, X. and Zhu, T.
TITLE      Stress-regulated genes of plants, transgenic plants containing
same, and methods of use

```


Search completed: May 26, 2004, 17:22:32
Job time : 349.756 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	42	100.0	42	15	US-10-222-952A-5	Sequence 5, Appli
2	42	100.0	401	13	US-10-282-122A-3339	Sequence 3339, Ap
3	40	95.2	40	15	US-10-222-952A-6	Sequence 6, Appli
4	25.6	61.0	90650	15	US-10-475-523-80	Sequence 80, Appli
5	25	59.5	25	15	US-10-222-952A-10	Sequence 9, Appli
6	25	59.5	25	15	US-10-222-952A-10	Sequence 10, Appli
7	24.8	59.0	1494	13	US-10-424-599-106547	Sequence 106547,
8	23.4	55.7	29956	12	US-09-997-722-229	Sequence 229, App
9	22.4	53.3	557	13	US-10-027-632-36277	Sequence 36277, A
10	22.4	53.3	557	13	US-10-027-632-36278	Sequence 36278, A
11	22.4	53.3	557	13	US-10-027-632-36279	Sequence 36279, A
12	22.4	53.3	557	13	US-10-027-632-62364	Sequence 62364, A
13	22.4	53.3	557	13	US-10-027-632-62365	Sequence 62365, A
14	22.4	53.3	557	13	US-10-027-632-62366	Sequence 62366, A

```

; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chilsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3339
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Escherichia coli
; US-10-282-122A-3339

Query Match 100.0%; Score 42; DB 13; Length 401;
Best Local Similarity 100.0%; Pred. No. 3.9e-05;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCATTACACATTCGTTAAGTCATATATGTTTTGACTTA 42
DB 145 CTGCATTACACATTCGTTAAGTCATATATGTTTTGACTTA 104

RESULT 3
US-10-222-952A-6/c
; Sequence 6, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regeneis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; PRIOR FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 40
; TYPE: DNA

```

```

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CHROML1B bottom, long oligo sequence
US-10-222-952A-6

Query Match 95.2%; Score 40; DB 15; Length 40;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GCACCTTACACATTCGTTAAGTCATATATGTTTTGACTTA 42
DB 40 GCACCTTACACATTCGTTAAGTCATATATGTTTTGACTTA 1

RESULT 4
US-10-175-523-80/c
; Sequence 80, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Palfreyman, Michael
; APPLICANT: Rajan, Priithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/1J795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 80
; LENGTH: 90650
; TYPE: DNA
; ORGANISM: Mus musculus
; US-10-175-523-80

Query Match 61.0%; Score 25.6; DB 15; Length 90650;
Best Local Similarity 77.5%; Pred. No. 88;
Matches 31; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CTGCATTACACATTCGTTAAGTCATATATGTTTTGACT 40
DB 66115 CTGCATTACATTCGTTTATACATATATTTGACT 56076

RESULT 5
US-10-222-952A-9
; Sequence 9, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regeneis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13

```

```

Query Match          59.0%; Score 24.8; DB 13; Length 1494;
Best Local Similarity 80.8%; Pred. No. 61;
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      7  TTACACATTCGGTTAAAGTCATATATGTTTTGACTTA 42
      | | | | | | | | | | | | | | | | | | | |
DB      1332 TAATTCATTCGCTAAATCATATATATTTTCGACTTA 1297

RESULT 8
US-09-997-722-229
; Sequence 229, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCES: A-71171/RMS/DCF
; CURRENT APPLICATION NUMBER: US/09/997,722
; PRIOR FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 229
; LENGTH: 29956
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: {1335}..(1858)
; OTHER INFORMATION: "n" at positions 1335 through 1858 can be any base.
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: {7091}..(7110)
; OTHER INFORMATION: "n" at positions 7091 through 7110 can be any base.
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: {20085}..(20586)
; OTHER INFORMATION: "n" at positions 20085 through 20586 can be any base.
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: {24372}..(24840)
; OTHER INFORMATION: "n" at positions 24372 through 24840 can be any base.
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: {29121}..(29589)
; OTHER INFORMATION: "n" at positions 29121 through 29589 can be any base.
US-09-997-722-229

Query Match          55.7%; Score 23.4; DB 12; Length 29956;
Best Local Similarity 73.2%; Pred. No. 4e+02;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY      1  CTGCACATTCACATCGTTGAAGTCATATATGTTTTGACTT 41
      | | | | | | | | | | | | | | | | | | | |
DB      18517 CTGCATTCATTCATTCATTCATTCATTCATTCATTCATTTTGATTT 18557

RESULT 9
US-10-027-632-36277/c
; Sequence 36277, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30

```


US-10-027-632-62964

Query Match 53.3%; Score 22.4; DB 13; Length 557;
Best Local Similarity 72.5%; Pred. No. 3.4e+02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 TGCACATTACACATTCGTTAAGTCATATATGTTTTGACTT 41
DB 490 TGCCCATAAACATTCCTTAAGAAGTATATACTATGACTT 529

RESULT 13

US-10-027-632-62965
; Sequence 62965, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62965
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-62965

Query Match 53.3%; Score 22.4; DB 13; Length 557;
Best Local Similarity 72.5%; Pred. No. 3.4e+02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 TGCACATTACACATTCGTTAAGTCATATATGTTTTGACTT 41
DB 490 TGCCCATAAACATTCCTTAAGAAGTATATACTATGACTT 529

RESULT 14

US-10-027-632-62966
; Sequence 62966, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62966
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-62966

Query Match 53.3%; Score 22.4; DB 13; Length 557;
Best Local Similarity 72.5%; Pred. No. 3.4e+02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 TGCACATTACACATTCGTTAAGTCATATATGTTTTGACTT 41
DB 490 TGCCCATAAACATTCCTTAAGAAGTATATACTATGACTT 529

RESULT 15

US-10-027-632-310230
; Sequence 310230, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 310230
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-310230

Query Match 53.3%; Score 22.4; DB 13; Length 557;
Best Local Similarity 72.5%; Pred. No. 3.4e+02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 TGCACATTACACATTCGTTAAGTCATATATGTTTTGACTT 41
DB 490 TGCCCATAAACATTCCTTAAGAAGTATATACTATGACTT 529

Search completed: May 27, 2004, 06:16:19
Job time : 864.256 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 17:50:29 ; Search time 821.197 Seconds
(without alignments)
221.574 Million cell updates/sec

Title: US-10-676-299-6

Perfect score: 40

Sequence: 1 taagcacaataatgatgacttaacgaatgtgaatgac 40

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2960401 seqs, 2274450654 residues

Total number of hits satisfying chosen parameters: 5920802

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:
- 2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:
- 6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:
- 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:
- 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:
- 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:
- 13: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:
- 14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:
- 15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:
- 16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:
- 17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:
- 18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:
- 19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	40	15	US-10-222-952A-6
2	40	100.0	42	15	US-10-222-952A-5
3	40	100.0	401	13	US-10-282-122A-3339
4	25	62.5	25	15	US-10-222-952A-9
5	25	62.5	25	15	US-10-222-952A-10
6	24.8	62.0	1494	13	US-10-424-599-106547
7	24.2	60.5	90650	15	US-10-175-523-80
8	22.2	55.5	611	13	US-10-027-632-216951
9	22.2	55.5	611	16	US-10-027-632-216951
10	21.6	54.5	19082	15	US-10-311-455-399
11	21.6	54.0	1125	13	US-10-027-632-117031
12	21.6	54.0	1125	13	US-10-027-632-117032
13	21.6	54.0	1125	13	US-10-027-632-117033
14	21.6	54.0	1125	16	US-10-027-632-117031

15	21.6	54.0	1125	16	US-10-027-632-117032	Sequence 117032,
16	21.6	54.0	1125	16	US-10-027-632-117033	Sequence 117033,
17	21.6	54.0	3287	13	US-10-027-632-115440	Sequence 115440,
18	21.6	54.0	3287	16	US-10-027-632-115440	Sequence 115440,
19	21.6	54.0	15667	15	US-10-311-455-399	Sequence 2119, App
20	21.6	54.0	17929	16	US-10-292-798-453	Sequence 453, App
21	21.6	54.0	41104	9	US-09-816-685-3	Sequence 3, Appli
22	21.6	54.0	41104	13	US-10-639-708-3	Sequence 3, Appli
23	21.4	53.5	557	13	US-10-027-632-36277	Sequence 36277, A
24	21.4	53.5	557	13	US-10-027-632-36278	Sequence 36278, A
25	21.4	53.5	557	13	US-10-027-632-36279	Sequence 36279, A
26	21.4	53.5	557	13	US-10-027-632-62964	Sequence 62964, A
27	21.4	53.5	557	13	US-10-027-632-62965	Sequence 62965, A
28	21.4	53.5	557	13	US-10-027-632-62966	Sequence 62966, A
29	21.4	53.5	557	13	US-10-027-632-310230	Sequence 310230,
30	21.4	53.5	557	13	US-10-027-632-310231	Sequence 310231,
31	21.4	53.5	557	13	US-10-027-632-310232	Sequence 310232,
32	21.4	53.5	557	16	US-10-027-632-36277	Sequence 36277, A
33	21.4	53.5	557	16	US-10-027-632-36278	Sequence 36278, A
34	21.4	53.5	557	16	US-10-027-632-36279	Sequence 36279, A
35	21.4	53.5	557	16	US-10-027-632-62964	Sequence 62964, A
36	21.4	53.5	557	16	US-10-027-632-62965	Sequence 62965, A
37	21.4	53.5	557	16	US-10-027-632-62966	Sequence 62966, A
38	21.4	53.5	557	16	US-10-027-632-310230	Sequence 310230,
39	21.4	53.5	557	16	US-10-027-632-310231	Sequence 310231,
40	21.4	53.5	557	16	US-10-027-632-310232	Sequence 310232,
41	21.4	53.5	627	13	US-10-027-632-3241	Sequence 3241, Ap
42	21.4	53.5	627	16	US-10-027-632-3241	Sequence 3241, Ap
43	21.4	53.5	1719	15	US-10-032-585-6736	Sequence 6736, Ap
44	21.4	53.5	29956	12	US-09-997-722-229	Sequence 229, App
45	21.2	53.0	1773	16	US-10-398-221-4010	Sequence 4010, Ap

ALIGNMENTS

RESULT 1
US-10-222-952A-6
; Sequence 6, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regensis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 40
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CHROML1B bottom, long oligo sequence
US-10-222-952A-6

Query Match 100.0%; Score 40; DB 15; Length 40;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAGTCAAAAACATATATGACTTAACGAATGTGAATGTC 40
|||||
DB 1 TAAGTCAAAAACATATATGACTTAACGAATGTGAATGTC 40

RESULT 2
US-10-222-952A-5/c
; Sequence 5, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:


```
; APPLICANT: Regensis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 42
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CHROMS1B top, long biotinylated oligo sequence; biotinylated
; OTHER INFORMATION: nucleotide at position 1
US-10-222-952A-5

Query Match      100.0%; Score 40; DB 15; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAGTCAAAACATATATGACTTAACGAATGTGTAAGTGC 40
Db 42 TAAGTCAAAACATATATGACTTAACGAATGTGTAAGTGC 3

RESULT 3
US-10-282-122A-3339
; Sequence 3339, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3339
```

```
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-282-122A-3339

Query Match      100.0%; Score 40; DB 13; Length 401;
Best Local Similarity 100.0%; Pred. No. 0.00002;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAGTCAAAACATATATGACTTAACGAATGTGTAAGTGC 40
Db 104 TAAGTCAAAACATATATGACTTAACGAATGTGTAAGTGC 143

RESULT 4
US-10-222-952A-9/c
; Sequence 9, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regensis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CHROMS1T top, short biotinylated oligo sequence; biotinylated
; OTHER INFORMATION: nucleotide at position 1
US-10-222-952A-9

Query Match      62.5%; Score 25; DB 15; Length 25;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAGTCAAAACATATATGACTTAA 25
Db 25 TAAGTCAAAACATATATGACTTAA 1

RESULT 5
US-10-222-952A-10
; Sequence 10, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regensis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CHROMS1B bottom short oligo sequence
US-10-222-952A-10

Query Match      62.5%; Score 25; DB 15; Length 25;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 TAAGTCAAAAACATATATGACTTAA 25
|||||
Db 1 TAAGTCAAAAACATATATGACTTAA 25

RESULT 6

US-10-424-599-106547
; Sequence 106547, Application US/10424599
; Publication No. US20040031072A1

GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 106547

; LENGTH: 1494

; TYPE: DNA

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_67228C.1

US-10-424-599-106547

Query Match 62.0%; Score 24.8; DB 13; Length 1494;
Best Local Similarity 80.6%; Pred. No. 58;
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 TAAGTCAAAAACATATATGACTTAA 36
|||||

Db 1297 TAAGTCAAAAACATATATGACTTAA 1332
|||||

RESULT 7

US-10-175-523-80

; Sequence 80, Application US/10175523

; Publication No. US20030096264A1

GENERAL INFORMATION:

; APPLICANT: Brockman, Jeffrey

; APPLICANT: Evans, David

; APPLICANT: Fook, Derek

; APPLICANT: Klimczak, Leszek

; APPLICANT: Laeng, Pascal

; APPLICANT: Palfreyman, Michael

; APPLICANT: Rajan, Prithi

; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)

; FILE REFERENCE: 3235/1J795-US3

; CURRENT APPLICATION NUMBER: US/10/175,523

; CURRENT FILING DATE: 2002-06-18

; PRIOR APPLICATION NUMBER: US 60/299,151

; PRIOR FILING DATE: 2001-06-18

; PRIOR APPLICATION NUMBER: US 60/317,828

; PRIOR FILING DATE: 2001-09-07

; PRIOR APPLICATION NUMBER: US 60/325,150

; PRIOR FILING DATE: 2001-09-25

; PRIOR APPLICATION NUMBER: US 60/333,047

; PRIOR FILING DATE: 2001-11-14

; PRIOR APPLICATION NUMBER: US 60/349,936

; PRIOR FILING DATE: 2002-01-18

; PRIOR APPLICATION NUMBER: US 60/361,834

; PRIOR FILING DATE: 2002-03-04

; NUMBER OF SEQ ID NOS: 197

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 80

; LENGTH: 90650

; TYPE: DNA

; ORGANISM: Mus musculus

US-10-175-523-80

Query Match 60.5%; Score 24.2; DB 15; Length 90650;
Best Local Similarity 78.4%; Pred. No. 2.4e+02;
Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 3 AGTCAAAACATATATGACTTAA 39
|||||

Db 66076 AGTCAAAACATATATGACTTAA 66112
|||||

RESULT 8

US-10-027-632-216951/c

; Sequence 216951, Application US/10027632

; Publication No. US20020198371A1

GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 216951

; LENGTH: 611

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-216951

Query Match 55.5%; Score 22.2; DB 13; Length 611;
Best Local Similarity 77.1%; Pred. No. 3.8e+02;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 AAGTCAAAAACATATATGACTTAA 36
|||||

Db 424 AAGTCAAAAACATATATGACTTAA 390
|||||

RESULT 9

US-10-027-632-216951/c

; Sequence 216951, Application US/10027632

; Publication No. US20030204075A9

GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

```
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 216951
; LENGTH: 611
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-216951

Query Match
Best Local Similarity 55.5%; Score 22.2; DB 16; Length 611;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 AAGTCAAAACATATATGACTTTAAGCAATGTGTAA 36
DB 424 AAGTCAGAAAATATATATAATAAAGACGCGGTAA 390

RESULT 10
US-10-311-455-599/c
; Sequence 599, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: BIERENSOCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 599
; LENGTH: 19082
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-599

Query Match
Best Local Similarity 54.5%; Score 21.8; DB 15; Length 19082;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 6 CAAAACATATATGACTTTAAGCAATGTGTAA 38
DB 7388 CAAAACATATATGACTTTAAGCAATGTGTAA 7356

RESULT 11
US-10-027-632-117031
; Sequence 117031, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117031
; LENGTH: 1125
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(1125)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-117031

Query Match
Best Local Similarity 54.0%; Score 21.6; DB 13; Length 1125;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 AAGTCAAAACATATATGACTTTAAGCAATGTGTAA 37
DB 1065 AATCACAAACAATATGGCTTCCAGATGTGTAA 1100

RESULT 12
US-10-027-632-117032
; Sequence 117032, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117032
; LENGTH: 1125
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(1125)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-117032

Query Match
Best Local Similarity 54.0%; Score 21.6; DB 13; Length 1125;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 AAGTCAAAACATATATGACTTTAAGCAATGTGTAA 37
```

Search completed: May 27, 2004, 06:16:19
Job time : 821.197 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 16:18:00 ; Search time 36.6624 Seconds
(without alignments)
348.146 Million cell updates/sec

Title: US-10-676-299-7

Perfect score: 23

Sequence: 1 ttaatcatatgcgttttggtta 23

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:*

- 1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/prodata/2/ina/FACTUS_COMB.seq:*
- 6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	18.8	81.7	269223	4	US-09-596-002-41
C 2	18.2	79.1	1026	4	US-09-134-001C-1995
C 3	17.8	77.4	1830121	4	US-09-557-884-1
C 4	17.8	77.4	1830121	4	US-09-643-990A-1
C 5	17.4	75.7	1386	4	US-09-543-681A-1519
C 6	17.2	74.8	627	4	US-09-669-751-43
C 7	16.8	73.0	678	4	US-09-543-681A-815
C 8	16.6	72.2	459	4	US-09-328-352-300
C 9	16.6	72.2	519	4	US-09-134-000C-724
C 10	16.6	72.2	2235	4	US-09-328-352-3794
C 11	16.6	72.2	7210	4	US-09-634-238-15
C 12	16.6	72.2	8537	4	US-10-204-708-41
C 13	16.6	72.2	640681	4	US-09-790-988-1
C 14	16.2	70.4	1442	4	US-09-227-357-119
C 15	16.2	70.4	1697	1	US-08-472-028A-9
C 16	16.2	70.4	1697	3	US-09-071-298-9
C 17	16.2	70.4	1697	3	US-09-196-268-9
C 18	16.2	70.4	1697	3	US-09-015-683-9
C 19	16.2	70.4	1697	4	US-09-191-998-9
C 20	16.2	70.4	580073	4	US-08-545-528D-1
C 21	16.2	70.4	1664976	4	US-08-916-421B-1
C 22	15.8	68.7	47	4	US-09-422-978-2280
C 23	15.8	68.7	1176	4	US-09-107-532A-765
C 24	15.8	68.7	1244	4	US-09-265-540S-3
C 25	15.8	68.7	202001	4	US-09-734-674-3
C 26	15.6	67.8	183	4	US-09-711-164-98
C 27	15.6	67.8	243	3	US-09-060-756-177

C 28	15.6	67.8	243	4	US-09-670-314-177	Sequence 177, App
C 29	15.6	67.8	336	3	US-09-060-756-174	Sequence 174, App
C 30	15.6	67.8	336	4	US-09-670-314-174	Sequence 174, App
C 31	15.6	67.8	467	3	US-09-060-756-662	Sequence 662, App
C 32	15.6	67.8	467	4	US-09-670-314-682	Sequence 662, App
C 33	15.6	67.8	678	4	US-09-328-352-2378	Sequence 2378, App
C 34	15.6	67.8	681	4	US-09-543-681A-17	Sequence 17, Appl
C 35	15.6	67.8	819	3	US-09-553-794-1	Sequence 1, Appl
C 36	15.6	67.8	1059	4	US-08-476-102A-8	Sequence 8, Appl
C 37	15.6	67.8	1626	4	US-09-540-236-1890	Sequence 1890, App
C 38	15.6	67.8	1698	4	US-09-134-000C-2368	Sequence 2368, App
C 39	15.6	67.8	1776	4	US-09-711-164-254	Sequence 254, App
C 40	15.6	67.8	1794	4	US-09-601-198-67	Sequence 67, Appl
C 41	15.6	67.8	1863	4	US-09-540-236-1189	Sequence 1189, App
C 42	15.6	67.8	2467	1	US-08-148-581D-1	Sequence 1, Appl
C 43	15.6	67.8	3001	4	US-09-539-333D-178	Sequence 178, App
C 44	15.6	67.8	3212	3	US-08-855-910-1	Sequence 1, Appl
C 45	15.6	67.8	6070	4	US-10-204-708-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-596-002-41/c
; Sequence 41, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 41
; LENGTH: 269223
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte template ID No. 6632636 41
; PUBLICATION INFORMATION:
US-09-596-002-41

Query Match 81.7%; Score 18.8; DB 4; Length 269223;
Best Local Similarity 90.9%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTAATCATATCGGTTTTGGTT 22
DB 142085 TTAATCATATCGGTTTTGGCT 142064

RESULT 2
US-09-134-001C-1995/c
; Sequence 1995, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 1995

; LENGTH: 1026

; TYPE: DNA

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-1995

Query Match 79.1%; Score 18.2; DB 4; Length 1026;

Best Local Similarity 87.0%; Pred. No. 14;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TTAATCATATGCGTTTGGTTA 23

Db 624 TAAATCATATGCGTTTGGTTA 602

RESULT 3

US-09-557-884-1/c

; Sequence 1, Application US/09557884

; Patent No. 6506581

; GENERAL INFORMATION:

APPLICANT: Fleischmann et al.

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/557,884

FILING DATE: 25-Apr-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/476,102

FILING DATE: JUN-5-1995

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB186P3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-309-8504

TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-557-884-1

Query Match

Best Local Similarity 77.4%; Score 17.8; DB 4; Length 1830121;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AATCATATGCGTTTGGTTA 23

Db 1238911 AACAAATATGCGTTTGGTTA 1238891

RESULT 4

US-09-643-990A-1/c

; Sequence 1, Application US/09643990A

; Patent No. 6528289

; GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann

Mark D. Adams

Owen White

Hamilton O. Smith

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville,

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,429

FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-610-5790

TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1

Query Match

Best Local Similarity 77.4%; Score 17.8; DB 4; Length 1830121;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AATCATATGCGTTTGGTTA 23

Db 1238911 AACAAATATGCGTTTGGTTA 1238891

RESULT 5

US-09-543-681A-1519

; Sequence 1519, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

APPLICANT: GARY BRETON

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRAB

TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 2709.1002-001

CURRENT APPLICATION NUMBER: US/09/543,681A

CURRENT FILING DATE: 2000-04-05

PRIOR APPLICATION NUMBER: US 60/128,706

PRIOR FILING DATE: 1999-04-09

NUMBER OF SEQ ID NOS: 8344

SEQ ID NO 1519

LENGTH: 1386

```
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-1519

Query Match      75.7%; Score 17.4; DB 4; Length 1386;
Best Local Similarity 94.7%; Pred. No. 33;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTGG 19
Db 422 TTAATCATATGCGTTTGG 440

RESULT 6
US-09-669-751-43/c
; Sequence 43, Application US/09669751
; Patent No. 6551575
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Ralph J.
; TITLE OF INVENTION: Methods for Identifying Compounds for
; TITLE OF INVENTION: Motion Sickness, Vertigo and Other Disorders Related to
; TITLE OF INVENTION: Balance and the Perception of Gravity
; FILE REFERENCE: P-NI 3864
; CURRENT APPLICATION NUMBER: US/09/669,751
; CURRENT FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 60/168,579
; PRIOR FILING DATE: 1999-12-02
; NUMBER OF SEQ ID NOS: 261
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Drosophila
US-09-669-751-43

Query Match      74.8%; Score 17.2; DB 4; Length 627;
Best Local Similarity 86.4%; Pred. No. 37;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTGGTT 22
Db 415 TTAATCAGAGCGTAATTTGGTT 394

RESULT 7
US-09-543-681A-815/c
; Sequence 815, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 815
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-815

Query Match      73.0%; Score 16.8; DB 4; Length 678;
Best Local Similarity 90.0%; Pred. No. 57;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTGG 20
Db 66 TAAATCATATGCTTTTGG 47

RESULT 8
US-09-328-352-300
; Sequence 300, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 300
; LENGTH: 459
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-300

Query Match      72.2%; Score 16.6; DB 4; Length 459;
Best Local Similarity 82.6%; Pred. No. 68;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTGGTTA 23
Db 72 TTATTCAGTTCGTTTGGTTA 94

RESULT 9
US-09-134-000C-724
; Sequence 724, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 724
; LENGTH: 519
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-724

Query Match      72.2%; Score 16.6; DB 4; Length 519;
Best Local Similarity 82.6%; Pred. No. 69;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTGGTTA 23
Db 40 TTGATTATATTTGTTTGGTTA 62

RESULT 10
US-09-328-352-3794/c
; Sequence 3794, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3794
; LENGTH: 2235
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
```

```
US-09-328-352-3794
Query Match          72.2%; Score 16.6; DB 4; Length 2235;
Best Local Similarity 82.6%; Pred. No. 79;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TTAATCATATGCGTTTGGTTA 23
Db 1214 TCAATCAATGCAATTTTGGTAA 1192

RESULT 11
US-09-634-238-15/c
; Sequence 15, Application US/09634238
; Patent No. 6544772
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka J.
; APPLICANT: Bloksberg, Leonard, N.
; APPLICANT: Lubbers, Mark W.
; APPLICANT: Dekker, James
; APPLICANT: Christensson, Anna C.
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul W.
; APPLICANT: Reid, Julian R.
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Polynucleotides, materials incorporating
; FILE REFERENCE: 11000.1043U1
; CURRENT APPLICATION NUMBER: US/09/634,238
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 7210
; TYPE: DNA
; ORGANISM: Lactobacillus rhamnosus
US-09-634-238-15

Query Match          72.2%; Score 16.6; DB 4; Length 7210;
Best Local Similarity 82.6%; Pred. No. 88;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TTAATCATATGCGTTTGGTTA 23
Db 4323 TAAGCATTTCGATTTTGGTTA 4301

RESULT 12
US-10-204-708-41
; Sequence 41, Application US/10204708
; Patent No. 667731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: BIERENBOCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; PRIOR FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 41

US-09-328-352-3794
; LENGTH: 8537
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-41

Query Match          72.2%; Score 16.6; DB 4; Length 8537;
Best Local Similarity 82.6%; Pred. No. 89;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TTAATCATATGCGTTTGGTTA 23
Db 7744 TTTAAATATTCGTTTGGTTA 7766

RESULT 13
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632335
; GENERAL INFORMATION:
; APPLICANT: SHIGEMOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match          72.2%; Score 16.6; DB 4; Length 640681;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TTAATCATATGCGTTTGGTTA 23
Db 516370 TTAATAATCTAAGTTTGGTTA 516348

RESULT 14
US-09-227-357-119
; Sequence 119, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
```


EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,932
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,916
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,930
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,918
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,920
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,733
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,795
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,919
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,928
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/055,722
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,723
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,948
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,949
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,953
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,950
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,947
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,964
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/056,360
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,684
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,984
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,954
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/058,785
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,664
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,660
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,661
EARLIER FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 672
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 119
LENGTH: 1442
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (1377)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1419)
OTHER INFORMATION: n equals a,t,g, or c
US-09-227-357-119

Query Match 70.4%; Score 16.2; DB 4; Length 1442;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 TAATCATATGCGTTTGGTT 22
|||||

DB 750 TAATCATATGTTTATGGTT 770
RESULT 15
US-08-472-028A-9
Sequence 9, Application US/08472028A
Patent No. 5767173
GENERAL INFORMATION:
APPLICANT: Ward, Eric R
TITLE OF INVENTION: Manipulation of Protoporphyrinogen
TITLE OF INVENTION: Oxidase Enzyme Activity in Eukaryotic Organisms
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,028A
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1748/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1697 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 29..1501
OTHER INFORMATION: /note= "Arabidopsis protox-3 cDNA;
OTHER INFORMATION: sequence from pMDC-5"
US-08-472-028A-9
Query Match 70.4%; Score 16.2; DB 1; Length 1697;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 3 AATCATATGCGTTTGGTTA 23
|||||
DB 1647 AATCATATGCGTTTGGTTA 1667
Search completed: May 26, 2004, 17:56:52
Job time : 43.6624 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 17:50:29 ; Search time 472.188 Seconds

(without alignments)
221.574 Million cell updates/sec

Title: US-10-676-299-7

Perfect score: 23

Sequence: 1 ttaatcatatgcgttttgggta 23

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 2960401 seqs, 2274450654 residues

Total number of hits satisfying chosen parameters: 5920802

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23	100.0	23	15	US-10-222-952A-7
2	23	100.0	30	15	US-10-222-952A-3
3	21	91.3	21	15	US-10-222-952A-8
4	21	91.3	28	15	US-10-222-952A-4
5	18.8	81.7	3727	16	US-10-094-749-622
6	18.8	81.7	269223	13	US-10-672-787-41
7	18.2	79.1	343	9	US-09-867-701-10855
8	17.8	77.4	505	13	US-10-424-599-90942
9	17.8	77.4	2475	16	US-10-383-630-1
10	17.8	77.4	10717	15	US-10-311-455-1667
11	17.8	77.4	1830121	15	US-10-329-960-1
12	17.8	77.4	1830121	16	US-10-329-670-1
13	17.4	75.7	1371	13	US-10-282-122A-32597
14	17.2	74.8	627	15	US-10-255-536-43

C 15	17.2	74.8	663	13	US-10-027-632-202848	Sequence 202848, A
C 16	17.2	74.8	663	13	US-10-027-632-202849	Sequence 202849, A
C 17	17.2	74.8	663	13	US-10-027-632-202850	Sequence 202850, A
C 18	17.2	74.8	663	16	US-10-027-632-202848	Sequence 202848, A
C 19	17.2	74.8	663	16	US-10-027-632-202849	Sequence 202849, A
C 20	17.2	74.8	663	16	US-10-027-632-202850	Sequence 202850, A
C 21	17.2	74.8	677	16	US-10-027-632-220322	Sequence 220322, A
C 22	17.2	74.8	677	16	US-10-027-632-220322	Sequence 220322, A
C 23	17.2	74.8	2000	9	US-09-938-842A-2707	Sequence 2707, Ap
C 24	17.2	74.8	2000	11	US-09-938-842A-2707	Sequence 2707, Ap
C 25	17.2	74.8	5216	15	US-10-311-455-319	Sequence 319, Appl
C 26	17.2	74.8	5216	15	US-10-240-485-11	Sequence 11, Appl
C 27	17.2	74.8	8423	13	US-10-221-613-401	Sequence 401, Appl
C 28	17.2	74.8	8423	15	US-10-311-455-1379	Sequence 1379, Ap
C 29	17.2	74.8	325348	16	US-10-085-117-358	Sequence 358, Appl
C 30	17.2	74.8	397658	9	US-09-813-320-3	Sequence 3, Appl
C 31	17.2	74.8	1601042	13	US-10-027-632-59064	Sequence 59064, A
C 32	17.2	74.8	1601042	16	US-10-027-632-59064	Sequence 59064, A
C 33	16.8	73.0	4169	15	US-10-369-493-34760	Sequence 34760, A
C 34	16.8	73.0	18154	15	US-10-311-455-228	Sequence 228, Appl
C 35	16.8	73.0	3673778	15	US-10-312-841-2	Sequence 2, Appl
C 36	16.6	72.2	177	13	US-10-424-599-110156	Sequence 110156, A
C 37	16.6	72.2	464	13	US-10-424-599-24604	Sequence 24604, A
C 38	16.6	72.2	501	13	US-10-027-632-64934	Sequence 64934, A
C 39	16.6	72.2	501	13	US-10-027-632-296688	Sequence 296688, A
C 40	16.6	72.2	501	16	US-10-027-632-64934	Sequence 64934, A
C 41	16.6	72.2	501	16	US-10-027-632-296688	Sequence 296688, A
C 42	16.6	72.2	503	13	US-10-027-632-37667	Sequence 37667, A
C 43	16.6	72.2	503	13	US-10-027-632-37667	Sequence 37667, A
C 44	16.6	72.2	533	13	US-10-027-632-42926	Sequence 42926, A
C 45	16.6	72.2	533	16	US-10-027-632-42926	Sequence 42926, A

ALIGNMENTS

RESULT 1

US-10-222-952A-7
; Sequence 7, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 23
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PLASMIT top, short biotinylated oligo sequence; biotinylated
; OTHER INFORMATION: nucleotide at position 1
US-10-222-952A-7

Query Match 100.0%; Score 23; DB 15; Length 23;
Best Local Similarity 100.0%; Pred.No. 2.8; Mismatches 0; Indels 0; Gaps 0;
Matches 23; Conservative 0;

QY 1 TTAATCATATGCGTTTGGTTA 23
Db 1 TTAATCATATGCGTTTGGTTA 23

RESULT 2

US-10-222-952A-3
; Sequence 3, Application US/10222952A
; Publication No. US20030096275A1

```
; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; PRIOR FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PLASLIT biotinylated top, long oligo sequence; biotinylated
; OTHER INFORMATION: nucleotide at position 1
US-10-222-952A-3

Query Match          91.3%; Score 23; DB 15; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTGGTTA 23
Db 1 TTAATCATATGCGTTTGGTTA 23

RESULT 3
US-10-222-952A-8/c
; Sequence 8, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PLASS1B bottom, short oligo sequence
US-10-222-952A-8

Query Match          91.3%; Score 21; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AATCATATGCGTTTGGTTA 23
Db 21 AATCATATGCGTTTGGTTA 1

RESULT 4
US-10-222-952A-4/c
; Sequence 4, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
```

```
; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; PRIOR FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 28
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PLASS1B bottom, long oligo sequence
US-10-222-952A-4

Query Match          91.3%; Score 21; DB 15; Length 28;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AATCATATGCGTTTGGTTA 23
Db 28 AATCATATGCGTTTGGTTA 8

RESULT 5
US-10-094-749-622/c
; Sequence 622, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 622
; LENGTH: 3727
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-622

Query Match          81.7%; Score 18.8; DB 16; Length 3727;
Best Local Similarity 90.9%; Pred. No. 3.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTGGTT 22
Db 2104 TTAATCATATGCGTTTGGTT 2083

RESULT 6
US-10-672-787-41/c
; Sequence 41, Application US/10672787
; Publication No. US20040067554A1
; GENERAL INFORMATION:
```

```
; APPLICANT: LAGACE, Robert, E.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BERG, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: ELIURA.025C1
; CURRENT APPLICATION NUMBER: US/10/672,787
; PRIOR FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: 09/596,002
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 41
; LENGTH: 269223
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-10-672-787-41

Query Match      81.7%; Score 18.8; DB 13; Length 269223;
Best Local Similarity 90.9%; Pred. No. 6.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TTAATCATATGCGTTTGGTT 22
      ||||| ||||| ||||| |||||
DB      142085 TTAATCATATGCGTTTGGCT 142064

RESULT 7
US-09-867-701-10855
; Sequence 10855, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10855
; LENGTH: 343
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-10855

Query Match      79.1%; Score 18.2; DB 9; Length 343;
Best Local Similarity 87.0%; Pred. No. 4.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TTAATCATATGCGTTTGGTTA 23
      ||||| ||||| ||||| |||||
DB      78 TTAATCAGATGCGCTTTTAGTTA 100

RESULT 8
US-10-424-599-90942
; Sequence 90942, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 90942
; LENGTH: 505
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
```

```
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MBT3847_53131C.1
US-10-424-599-90942

Query Match      77.4%; Score 17.8; DB 13; Length 505;
Best Local Similarity 90.5%; Pred. No. 6.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 TTAATCATATGCGTTTGGTT 22
      ||||| ||||| ||||| |||||
DB      299 TAAACATATGCGTTTGGTT 319

RESULT 9
US-10-383-630-1/c
; Sequence 1, Application US/10383630
; Publication No. US20040002431A1
; GENERAL INFORMATION:
; APPLICANT: HAKAMADA, YOSHITIRO
; APPLICANT: SAWADA, KAZUHISA
; APPLICANT: ENDO, KEIJI
; APPLICANT: KODAMA, HIROSHI
; APPLICANT: WADA, YASUNAO
; APPLICANT: SHIKATA, SHITSU
; APPLICANT: KOBAYASHI, TOHRU
; TITLE OF INVENTION: Mutant alkali cellulase
; FILE REFERENCE: 234890US0
; CURRENT APPLICATION NUMBER: US/10/383,630
; CURRENT FILING DATE: 2003-03-10
; PRIOR APPLICATION NUMBER: JP P2002-089531
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: JP P2003-013840
; PRIOR FILING DATE: 2003-01-22
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2475
; TYPE: DNA
; ORGANISM: Bacillus sp.KSM-S237
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2475)
; OTHER INFORMATION:
US-10-383-630-1

Query Match      77.4%; Score 17.8; DB 16; Length 2475;
Best Local Similarity 90.5%; Pred. No. 8.8e-02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TTAATCATATGCGTTTGGT 21
      ||||| ||||| ||||| |||||
DB      2102 TTAATCGTATACGTTTGGT 2082

RESULT 10
US-10-311-455-1667
; Sequence 1667, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
```

```
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1667
; LENGTH: 10717
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1667

Query Match          77.4%; Score 17.8; DB 15; Length 10717;
Best Local Similarity 90.5%; Pred. No. 1.1e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TTAATCATATGCGTTTTCGT 21
DB 4757 TTAATTAGATGCGTTTTCGT 4777

RESULT 11
US-10-329-960-1/c
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
```

```
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a, t, g or c
```

```
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152530)..(152530)
Query Match 77.4%; Score 17.8; DB 15; Length 1830121;
Best Local Similarity 90.5%; Pred. No. 1.5e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 AATCATATGCGTTTGGTTA 23
Db 1238911 AACATATGCGTTTGGTTA 1238891
RESULT 12
US-10-323-670-1/c
; Sequence 1, Application US/10329670
; Publication No. US20040018503A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fra
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186PL
; CURRENT APPLICATION NUMBER: US/10/329,670
; CURRENT FILING DATE: 2002-12-24
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
```

```

; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature

```

```

; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152530)..(152530)

```

Query Match 77.4%; Score 17.8; DB 16; Length 1830121;
 Best Local Similarity 90.5%; Pred. No. 1.5e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy 3 AATCATATGCGTTTTCGTTA 23
Db 1238911 AACATATGCGTTTTCGTTA 1238891

RESULT 13
US-10-282-122A-32597
; Sequence 32597, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Orlsen, Kari
; APPLICANT: Zvekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32597
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-10-282-122A-32597

Query Match 75.7%; Score 17.4; DB 13; Length 1371;
Best Local Similarity 94.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTAATCATATGCGTTTTCGTTG 19
Db 410 TTAATCCTATGCGTTTTCGTTG 428

RESULT 14
US-10-255-536-43/c
; Sequence 43, Application US/10255536
; Publication No. US20030087807A1
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Ralph J.
; TITLE OF INVENTION: Methods for Identifying Compounds for
; Motion Sickness, Vertigo and Other Disorders Related to
```

```
; TITLE OF INVENTION: Balance and the Perception of Gravity
; FILE REFERENCE: P-NI 3864
; CURRENT APPLICATION NUMBER: US/10/255,536
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: US/09/669,751
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 60/168,579
; PRIOR FILING DATE: 1999-12-02
; NUMBER OF SEQ ID NOS: 261
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Drosophila
US-10-255-536-43

Query Match 74.8%; Score 17.2; DB 15; Length 627;
Best Local Similarity 86.4%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TTAATCATATGCGTTTTCGTTT 22
Db 415 TTAATCAGAAGCGTATTGTT 394

RESULT 15
US-10-027-632-202848/c
; Sequence 202848, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108927.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 202848
; LENGTH: 663
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(663)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-202848

Query Match 74.8%; Score 17.2; DB 13; Length 663;
Best Local Similarity 82.8%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TTAATCATATGCGTTTTCGTTA 23
Db 46 TTAATCATATGCGTTTTCGTTA 24

Search completed: May 27, 2004, 06:16:38
Job time : 491.188 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 17:50:29 ; Search time 431.128 Seconds
(without alignments)
221.574 Million cell updates/sec

Title: US-10-676-299-8

Perfect score: 21

Sequence: 1 taacacaaacgcataatgatt 21

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2960401 seqs, 2274450654 residues

Total number of hits satisfying chosen parameters: 5920802

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCT05_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US50_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US50_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	15	US-10-222-952A-8
2	21	100.0	23	15	US-10-222-952A-7
3	21	100.0	28	15	US-10-222-952A-4
4	21	100.0	30	15	US-10-222-952A-3
5	17.8	84.8	1830121	15	US-10-329-960-1
6	17.8	84.8	1830121	16	US-10-329-670-1
7	17.4	82.9	3727	16	US-10-094-749-622
8	16.8	80.0	505	13	US-10-424-599-90942
9	16.8	80.0	4169	16	US-10-369-493-34760
10	16.8	80.0	18154	13	US-10-311-455-228
11	16.8	80.0	269223	13	US-10-672-787-41
12	16.8	80.0	367378	15	US-10-312-841-2
13	16.4	78.1	33805	13	US-10-087-182-1483
14	16.2	77.1	177	13	US-10-424-599-110156

C 15	16.2	77.1	343	9	US-09-867-701-10855	Sequence 10855, A
C 16	16.2	77.1	663	13	US-10-027-632-218409	Sequence 218409, A
C 17	16.2	77.1	663	16	US-10-027-632-218409	Sequence 218409, A
C 18	16.2	77.1	1942	13	US-10-425-114-16056	Sequence 16056, A
C 19	16.2	77.1	2000	9	US-09-938-842A-3670	Sequence 3670, A
C 20	16.2	77.1	2000	9	US-09-938-842A-5323	Sequence 5323, A
C 21	16.2	77.1	2000	11	US-09-938-842A-3670	Sequence 3670, A
C 22	16.2	77.1	2000	11	US-09-938-842A-5323	Sequence 5323, A
C 23	16.2	77.1	2878	16	US-10-108-260A-1168	Sequence 1168, A
C 24	16.2	77.1	7210	16	US-10-264-213-15	Sequence 15, Appl
C 25	16.2	77.1	8423	15	US-10-311-455-1379	Sequence 1379, A
C 26	16.2	77.1	79860	13	US-10-087-192-412	Sequence 412, App
C 27	16.2	77.1	325348	16	US-10-085-117-358	Sequence 358, App
C 28	16.2	77.1	580073	15	US-10-205-220-1	Sequence 1, Appl
C 29	16.2	77.1	2940917	13	US-10-027-632-174763	Sequence 174763, A
C 30	16.2	77.1	2940917	16	US-10-027-632-174763	Sequence 174763, A
C 31	16	76.2	2814	15	US-10-311-455-2100	Sequence 2100, A
C 32	15.8	75.2	47	16	US-10-349-143-2280	Sequence 2280, A
C 33	15.8	75.2	594	10	US-09-764-891-2095	Sequence 2095, A
C 34	15.8	75.2	594	15	US-10-091-572-126	Sequence 126, App
C 35	15.8	75.2	646	13	US-10-027-632-97093	Sequence 97093, A
C 36	15.8	75.2	646	13	US-10-027-632-306448	Sequence 306448, A
C 37	15.8	75.2	646	16	US-10-027-632-97093	Sequence 97093, A
C 38	15.8	75.2	646	16	US-10-027-632-306448	Sequence 306448, A
C 39	15.8	75.2	650	13	US-10-027-632-185434	Sequence 185434, A
C 40	15.8	75.2	650	13	US-10-027-632-185435	Sequence 185435, A
C 41	15.8	75.2	650	16	US-10-027-632-185434	Sequence 185434, A
C 42	15.8	75.2	650	16	US-10-027-632-185435	Sequence 185435, A
C 43	15.8	75.2	781	13	US-10-424-599-96122	Sequence 96122, A
C 44	15.8	75.2	895	10	US-09-764-891-9170	Sequence 9170, A
C 45	15.8	75.2	895	15	US-10-091-572-746	Sequence 746, App

ALIGNMENTS

RESULT 1
US-10-222-952A-8
; Sequence 8, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/IL443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PLASS1B bottom, short oligo sequence
US-10-222-952A-8

Query Match 100.0%; Score 21; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAACCAAAACGCATATGATT 21
DB 1 TAACCAAAACGCATATGATT 21
RESULT 2
US-10-222-952A-7/c
; Sequence 7, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:

```
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 23
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PLAS11T top, short biotinylated oligo sequence; biotinylated
; OTHER INFORMATION: nucleotide at position 1
US-10-222-952A-7
```

```
Query Match 100.0%; Score 21; DB 15; Length 23;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TAACCAAAAACGCATATGATT 21
   |||||
Db 23 TAACCAAAAACGCATATGATT 3
```

```
RESULT 3
US-10-222-952A-4
; Sequence 4, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 28
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PLAS11B bottom, long oligo sequence
US-10-222-952A-4
```

```
Query Match 100.0%; Score 21; DB 15; Length 28;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TAACCAAAAACGCATATGATT 21
   |||||
Db 8 TAACCAAAAACGCATATGATT 28
```

```
RESULT 4
US-10-222-952A-3/c
; Sequence 3, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
```

```
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PLAS11T biotinylated top, long oligo sequence; biotinylated
; OTHER INFORMATION: nucleotide at position 1
US-10-222-952A-3
```

```
Query Match 100.0%; Score 21; DB 15; Length 30;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TAACCAAAAACGCATATGATT 21
   |||||
Db 23 TAACCAAAAACGCATATGATT 3
```

```
RESULT 5
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fr
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1930121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29258)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
```

```

; LOCATION: {40808}..{40810}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {44416}..{44416}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {44905}..{44905}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {44975}..{44975}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {45593}..{45593}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {45732}..{45732}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {47036}..{47036}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {51334}..{51334}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {51602}..{51602}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {51786}..{51786}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {51805}..{51805}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {55369}..{55369}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {65309}..{65309}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {65313}..{65313}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {80024}..{80024}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {100091}..{100091}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {102696}..{102696}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {105121}..{105121}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {107248}..{107248}

```

```

; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {117136}..{117136}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {119750}..{119750}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {119924}..{119924}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {120038}..{120038}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {121344}..{121344}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {122167}..{122167}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {122336}..{122336}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {131340}..{131340}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {131360}..{131360}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {139910}..{139910}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {140398}..{140398}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {142750}..{142750}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {145058}..{145058}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {145171}..{145171}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {145942}..{145942}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {147197}..{147197}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {150841}..{150841}
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: {152500}..{152500}
; OTHER INFORMATION: n equals a, t, g or c

```


LOCATION: (119750)..(119750)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (119924)..(119924)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (120038)..(120038)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (121344)..(121344)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (122167)..(122167)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (122336)..(122336)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (131340)..(131340)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (131360)..(131360)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (139910)..(139910)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (140398)..(140398)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (142750)..(142750)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (145058)..(145058)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (145171)..(145171)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (145942)..(145942)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (147197)..(147197)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (150841)..(150841)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (152500)..(152500)
 OTHER INFORMATION: n equals a, t, g or c
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (152530)..(152530)

Query Match 84.8%; Score 17.8; DB 16; Length 1830121;
 Best Local Similarity 90.5%; Pred. No. 7.6e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TAACCAAAACGCATATGAT 21
 Db 1238891 TAACCAAAACGCATATGTT 1238911

RESULT 7

US-10-094-749-622
 ; Sequence 622, Application US/10094749
 ; Publication No. US20030219741A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ISOGAI, TAKAO
 ; APPLICANT: SUGIYAMA, TOMOYASU
 ; APPLICANT: OTSUKI, TETSUJI
 ; APPLICANT: WAKAMATSU, AI
 ; APPLICANT: SATO, HIROYUKI
 ; APPLICANT: ISHII, SHIZUKO
 ; APPLICANT: YAMAMOTO, JUN-ICHI
 ; APPLICANT: ISONO, YUUKO
 ; APPLICANT: HIO, YURI
 ; APPLICANT: OTSUKA, KAORU
 ; APPLICANT: NAGAI, KEIICHI
 ; APPLICANT: IRIE, RYOTARO
 ; APPLICANT: TAMECHIKA, ICHIRO
 ; APPLICANT: SEKI, NACHIKO
 ; APPLICANT: YOSHIKAWA, TSUTOMU
 ; APPLICANT: OTSUKA, MOTOYUKI
 ; APPLICANT: NAGAHARI, KENJI
 ; APPLICANT: MASUHO, YASUHIKO
 ; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
 ; FILE REFERENCE: 084335/0160
 ; CURRENT APPLICATION NUMBER: US/10/094,749
 ; CURRENT FILING DATE: 2002-03-12
 ; PRIOR APPLICATION NUMBER: 60/350,435
 ; PRIOR FILING DATE: 2002-01-24
 ; PRIOR APPLICATION NUMBER: JP 2001-328381
 ; PRIOR FILING DATE: 2001-09-14
 ; NUMBER OF SEQ ID NOS: 3381
 ; SOFTWARE: Patent in Ver. 2.1
 ; SEQ ID NO 622
 ; LENGTH: 3727
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-094-749-622

Query Match 82.9%; Score 17.4; DB 15; Length 3727;
 Best Local Similarity 94.7%; Pred. No. 5.2e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 AACCAAAACGCATATGAT 20
 Db 2083 AACCAAAACGCATATGAT 2101

RESULT 8

US-10-424-599-90942/c
 ; Sequence 90942, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 90942
 ; LENGTH: 505
 ; TYPE: DNA
 ; ORGANISM: Glycine max

FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_53131C.1
US-10-424-599-90942

Query Match 80.0%; Score 16.8; DB 13; Length 505;
Best Local Similarity 90.0%; Pred. No. 7.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TAACCAAAACGCATATGATT 21
|||||
Db 319 AACCAAAACGCATATGTTT 300
|||||

RESULT 9
US-10-369-493-34760
; Sequence 34760, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 34760
; LENGTH: 4169
; TYPE: DNA
; ORGANISM: Ferrioplasma acidarmanus
US-10-369-493-34760

Query Match 80.0%; Score 16.8; DB 16; Length 4169;
Best Local Similarity 90.0%; Pred. No. 9.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAACCAAAACGCATATGAT 20
|||||
Db 2074 TAACCAAAACACATATGCT 2093
|||||

RESULT 10
US-10-311-455-228/c
; Sequence 228, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 228
; LENGTH: 18154
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-228

Query Match 80.0%; Score 16.8; DB 15; Length 18154;
Best Local Similarity 90.0%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAACCAAAACGCATATGAT 20
|||||
Db 9240 TAACCAAAACGCATATAAT 9221
|||||

RESULT 11
US-10-672-787-41
; Sequence 41, Application US/10672787
; Publication No. US20040067554A1
; GENERAL INFORMATION:
; APPLICANT: LAGACE, Robert, E.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BERG, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: ELITRA.025C1
; CURRENT APPLICATION NUMBER: US/10/672,787
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: 09/596,002
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 41
; LENGTH: 269223
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-10-672-787-41

Query Match 80.0%; Score 16.8; DB 13; Length 269223;
Best Local Similarity 90.0%; Pred. No. 1.8e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AACCAAAACGCATATGATT 21
|||||
Db 142064 AGCCAAACGCATATCATT 142083
|||||

RESULT 12
US-10-312-841-2/c
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des M
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2 3673778
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (379615)
US-10-312-841-2

Query Match 80.0%; Score 16.8; DB 15; Length 3673778;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AACCAAAACGCATATGATT 21
|||||
Db 3080035 AACCTAAACGCATATATT 3080016
|||||

RESULT 13

US-10-087-192-1483
; Sequence 1483, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1483
; LENGTH: 33805
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(33805)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1483

Query Match 78.1%; Score 16.4; DB 13; Length 33805;
Best Local Similarity 94.4%; Pred. No. 2.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAACCAAAAACGCATATG 18
|||||
DB 653 TAACCAAAAACGCATATG 670

RESULT 14
US-10-424-599-110156/c
; Sequence 110156, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yitua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 110156
; LENGTH: 177
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(177)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_70483C.1
US-10-424-599-110156

Query Match 77.1%; Score 16.2; DB 13; Length 177;
Best Local Similarity 85.7%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAACCAAAAACGCATATG 21
|||||
DB 95 TAACCAAAAACGCATATG 75

RESULT 15

US-09-867-701-10855/c
; Sequence 10855, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10855
; LENGTH: 343
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-10855

Query Match 77.1%; Score 16.2; DB 9; Length 343;
Best Local Similarity 85.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAACCAAAAACGCATATG 21
|||||
DB 100 TAACCAAAAACGCATATG 80

Search completed: May 27, 2004, 06:17:00
Job time : 453.128 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 17:50:29 ; Search time 513.248 Seconds
(without alignments)
221.574 Million cell updates/sec

Title: US-10-676-299-9

Perfect score: 25

Sequence: 1 ttaagtcataatgttttgactta 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2960401 seqs, 2274450654 residues

Total number of hits satisfying chosen parameters: 5920802

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:

1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:
10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:
14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	15	US-10-222-952A-9
2	25	100.0	25	15	US-10-222-952A-10
3	25	100.0	40	15	US-10-222-952A-6
4	25	100.0	42	15	US-10-222-952A-5
5	25	100.0	401	13	US-10-282-122A-3339
6	20.8	83.2	650	13	US-10-027-632-197612
7	20.8	83.2	650	16	US-10-027-632-197612
8	20.2	80.8	2000	9	US-09-938-842A-3647
9	20.2	80.8	2000	11	US-09-938-842A-3647
10	20.2	80.8	5641	15	US-10-311-455-1370
11	19.8	79.2	798	13	US-10-424-599-135777
12	19.2	76.8	283	9	US-09-783-590-2768
13	19.2	76.8	537	13	US-10-027-632-206167
14	19.2	76.8	537	16	US-10-027-632-206167

C 15	19.2	76.8	1494	13	US-10-424-599-106547
C 16	18.8	75.2	640681	9	US-09-790-988-1
C 17	18.6	74.4	344	13	US-10-424-599-83711
C 18	18.6	74.4	366	13	US-10-424-599-80042
C 19	18.6	74.4	2274	13	US-10-027-632-262525
C 20	18.6	74.4	2274	16	US-10-027-632-262525
C 21	18.6	74.4	2478	13	US-10-424-599-126150
C 22	18.6	74.4	2691	16	US-10-369-493-46160
C 23	18.6	74.4	2753	13	US-10-424-599-49551
C 24	18.6	74.4	10279	15	US-10-311-455-1564
C 25	18.6	74.4	10279	17	US-10-240-589C-86
C 26	18.6	74.4	73334	15	US-10-311-455-2097
C 27	18.6	74.4	73334	17	US-10-240-589C-127
C 28	18.6	74.4	248436	13	US-10-087-192-2014
C 29	18.4	73.6	10891	15	US-10-311-455-438
C 30	18.2	72.8	480	13	US-10-424-599-56865
C 31	18.2	72.8	1486	13	US-10-115-123-73
C 32	18.2	72.8	1486	15	US-10-012-542-73
C 33	18.2	72.8	2121	15	US-10-006-852-7
C 34	18.2	72.8	10467	15	US-10-204-708-1
C 35	18.2	72.8	10672	10	US-09-784-891-5870
C 36	18.2	72.8	60452	13	US-10-087-192-1900
C 37	18.2	72.8	118931	13	US-10-087-192-1108
C 38	18.2	72.8	358246	16	US-10-292-798-1095
C 39	17.8	71.2	718	13	US-10-027-632-164169
C 40	17.8	71.2	718	16	US-10-027-632-164169
C 41	17.8	71.2	2125	13	US-10-425-114-2142
C 42	17.8	71.2	193303	15	US-10-081-327-37
C 43	17.8	71.2	193303	15	US-10-081-327-44
C 44	17.8	71.2	1163020	16	US-10-398-221-10
C 45	17.8	71.2	3011208	16	US-10-398-221-2058

ALIGNMENTS

RESULT 1

US-10-222-952A-9

; Sequence 9, Application US/10222952A

; Publication No. US20030096275A1

; GENERAL INFORMATION:

; APPLICANT: Regensis

; APPLICANT: Laing, Lance

; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES

; FILE REFERENCE: 4107/11443-US1

; CURRENT APPLICATION NUMBER: US/10/222,952A

; CURRENT FILING DATE: 2002-08-15

; PRIOR APPLICATION NUMBER: US 60/313,714

; PRIOR FILING DATE: 2001-08-20

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 9

; LENGTH: 25

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: CHROMSIT top, short biotinylated oligo sequence; biotinylated
; OTHER INFORMATION: nucleotide at position 1

US-10-222-952A-9

Query Match 100.0%; Score 25; DB 15; Length 25;
Best Local Similarity 100.0%; Pred No. 2,2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTTGACTTA 25

Db 1 TTAAGTCATATATGTTTTGACTTA 25

RESULT 2

US-10-222-952A-10/c

; Sequence 10, Application US/10222952A

; Publication No. US20030096275A1

```
; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/1L443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; PRIOR FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CHROMS1B bottom short oligo sequence
; US-10-222-952A-10

Query Match      100.0%; Score 25; DB 15; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTAAGTCATATATGTTTTTGACTTA 25
Db 25 TTAAGTCATATATGTTTTTGACTTA 1

RESULT 3
US-10-222-952A-6/c
; Sequence 6, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/1L443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; PRIOR FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 40
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CHROML1B bottom, long oligo sequence
; US-10-222-952A-6

Query Match      100.0%; Score 25; DB 15; Length 40;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTAAGTCATATATGTTTTTGACTTA 25
Db 25 TTAAGTCATATATGTTTTTGACTTA 1

RESULT 4
US-10-222-952A-5
; Sequence 5, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/1L443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; PRIOR FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3339
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Escherichia coli
; US-10-282-122A-3339

Query Match      100.0%; Score 25; DB 13; Length 401;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/1L443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; PRIOR FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 42
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CHROML1T top, long biotinylated oligo sequence; biotinylated
; US-10-222-952A-5

Query Match      100.0%; Score 25; DB 15; Length 42;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTAAGTCATATATGTTTTTGACTTA 25
Db 18 TTAAGTCATATATGTTTTTGACTTA 42

RESULT 5
US-10-282-122A-3339/c
; Sequence 3339, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3339
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Escherichia coli
; US-10-282-122A-3339

Query Match      100.0%; Score 25; DB 13; Length 401;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 TTAAGTCATATATGTTTTGACTTA 25
Db 128 TTAAGTCATATATGTTTTGACTTA 104

RESULT 6
US-10-027-632-197612/c
; Sequence 197612, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 197612
; LENGTH: 650
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-197612

Query Match 83.2%; Score 20.8; DB 13; Length 650;
Best Local Similarity 91.7%; Pred. No. 1.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TTAAGTCATATATGTTTTGACTT 24
Db 492 TTAAGTCATATATGTTTTAGAGTT 469

RESULT 7
US-10-027-632-197612/c
; Sequence 197612, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09

Qy 1 TTAAGTCATATATGTTTTGACTT 24
Db 492 TTAAGTCATATATGTTTTAGAGTT 469

RESULT 8
US-09-938-842A-3647/c
; Sequence 3647, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3647
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3647

Query Match 80.8%; Score 20.2; DB 9; Length 2000;
Best Local Similarity 88.0%; Pred. No. 3.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TTAAGTCATATATGTTTTGACTTA 25
Db 975 TTAAGTCATATATGTTTTAGATT 951

RESULT 9
US-09-938-842A-3647/c
; Sequence 3647, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
```

```

; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3647
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3647

Query Match      80.8%; Score 20.2; DB 11; Length 2000;
Best Local Similarity 88.0%; Pred No. 3.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTGGACTTA 25
   |||||
Db 975 TTATGTCATATATGTTTGGACTTA 951

RESULT 10
US-10-311-455-1370
; Sequence 1370, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Cytosine Methylation of Cytosine
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1370
; LENGTH: 5641
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1370

Query Match      80.8%; Score 20.2; DB 15; Length 5641;
Best Local Similarity 88.0%; Pred No. 3.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTGGACTTA 25
   |||||
Db 1512 TTATGTCATATATGTTTGGACTTA 1536

RESULT 11
US-10-424-599-135777/c
; Sequence 135777, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David X
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 28584
; SEQ ID NO 135777
; LENGTH: 798
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:

```

```

; OTHER INFORMATION: Clone ID: PAT_MRT3847_93615C.1
US-10-424-599-135777

Query Match      79.2%; Score 19.8; DB 13; Length 798;
Best Local Similarity 91.3%; Pred No. 4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TTAAGTCATATATGTTTGGACTT 24
   |||||
Db 221 TTAAGTCATATATGTTTGGACTT 199

RESULT 12
US-09-783-590-2768/c
; Sequence 2768, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PQ-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2768
; LENGTH: 283
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (140)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (179)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (205)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (206)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (248)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-2768

Query Match      76.8%; Score 19.2; DB 9; Length 283;
Best Local Similarity 84.0%; Pred No. 5.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTGGACTTA 25
   |||||
Db 145 TTAAGTCATATATGTTTGGACTTA 121

RESULT 13
US-10-027-632-206167/c
; Sequence 206167, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632

```

; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 206167
; LENGTH: 537
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(537)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-206167

Query Match 76.8%; Score 19.2; DB 13; Length 537;
Best Local Similarity 87.5%; Pred. No. 6.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TTAAGTCATATATGTTTGGACTT 24
Db 263 TTAAGTCATATATGTTTGGACTT 240

RESULT 14
US-10-027-632-206167/c
; Sequence 206167, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 206167
; LENGTH: 537
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(537)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-206167

Query Match 76.8%; Score 19.2; DB 16; Length 537;
Best Local Similarity 87.5%; Pred. No. 6.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TTAAGTCATATATGTTTGGACTT 24
Db 263 TTAAGTCATATATGTTTGGACTT 240

RESULT 15
US-10-424-599-106547/c
; Sequence 106547, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Zhou Yihua
; APPLICANT: Kovalic David K
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 106547
; LENGTH: 1494
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_67228C.1
US-10-424-599-106547

Query Match 76.8%; Score 19.2; DB 13; Length 1494;
Best Local Similarity 87.5%; Pred. No. 7.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 TTAAGTCATATATGTTTGGACTT 25
Db 1320 TTAAGTCATATATGTTTGGACTT 1297

Search completed: May 27, 2004, 06:17:06
Job time : 519.248 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 17:50:29 ; Search time 513.248 Seconds
(without alignments)
221.574 Million cell updates/sec

Title: US-10-676-299-10

Perfect score: 25
Sequence: 1 taagtcacaaacatatatgacttaa 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2960401 seqs, 2274450654 residues

Total number of hits satisfying chosen parameters: 5920802

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
C 1	25	100.0	25	15	US-10-222-952A-9
2	25	100.0	25	15	US-10-222-952A-10
3	25	100.0	40	15	US-10-222-952A-6
C 4	25	100.0	42	15	US-10-222-952A-5
5	25	100.0	401	13	US-10-282-122A-3339
6	20.8	83.2	650	13	US-10-027-632-197612
7	20.8	83.2	650	16	US-10-027-632-197612
8	20.2	80.8	2000	9	US-09-938-842A-3647
9	20.2	80.8	2000	11	US-09-938-842A-3647
C 10	20.2	80.8	5641	15	US-10-311-455-1370
11	19.8	79.2	798	13	US-10-424-599-135777
12	19.2	76.8	283	9	US-09-783-590-2768
C 13	19.2	76.8	537	13	US-10-027-632-206167
14	19.2	76.8	537	16	US-10-027-632-206167

15	19.2	76.8	1494	13	US-10-424-599-106547
16	18.8	75.2	640681	9	US-09-790-988-1
C 17	18.6	74.4	344	13	US-10-424-599-83711
18	18.6	74.4	366	13	US-10-424-599-80042
C 19	18.6	74.4	2274	13	US-10-027-632-262525
C 20	18.6	74.4	2274	16	US-10-027-632-262525
C 21	18.6	74.4	2478	13	US-10-424-599-126150
C 22	18.6	74.4	2691	16	US-10-369-493-46160
C 23	18.6	74.4	2753	13	US-10-424-599-49551
C 24	18.6	74.4	10279	15	US-10-311-455-1564
C 25	18.6	74.4	10279	17	US-10-240-589C-86
C 26	18.6	74.4	73334	15	US-10-311-455-2097
C 27	18.6	74.4	73334	17	US-10-240-589C-127
28	18.6	74.4	248436	13	US-10-087-192-2014
C 29	18.4	73.6	10891	15	US-10-311-455-438
C 30	18.2	72.8	480	13	US-10-424-599-5685
C 31	18.2	72.8	1486	13	US-10-115-123-73
C 32	18.2	72.8	1486	15	US-10-012-542-73
C 33	18.2	72.8	2121	15	US-10-006-852-7
C 34	18.2	72.8	10457	15	US-10-204-708-1
C 35	18.2	72.8	10872	10	US-09-764-891-5870
C 36	18.2	72.8	60452	13	US-10-087-192-1900
C 37	18.2	72.8	118931	13	US-10-087-192-1108
C 38	18.2	72.8	358246	16	US-10-292-798-1095
C 39	17.8	71.2	718	13	US-10-027-632-164169
C 40	17.8	71.2	718	16	US-10-027-632-164169
C 41	17.8	71.2	2125	13	US-10-425-114-2142
C 42	17.8	71.2	193303	15	US-10-081-327-37
C 43	17.8	71.2	193303	15	US-10-081-327-44
C 44	17.8	71.2	1163020	16	US-10-398-221-10
45	17.8	71.2	3011208	16	US-10-398-221-2058

ALIGNMENTS

RESULT 1

US-10-222-952A-9/c
; Sequence 9, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CHROMSIT top, short biotinylated oligo sequence; biotinylated nucleotide at position 1
US-10-222-952A-9

Query Match 100.0%; Score 25; DB 15; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.2; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0;
QY 1 TAAGTCAAAACATATATGACTTAA 25
|||||
DB 25 TAAGTCAAAACATATATGACTTAA 1

RESULT 2

US-10-222-952A-10
; Sequence 10, Application US/10222952A
; Publication No. US20030096275A1

```

; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; PRIOR FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CHROMS1B bottom short oligo sequence
US-10-222-952A-10

Query Match      100.0%; Score 25; DB 15; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TAAGTCAAAAACATATATGACTTAA 25
Db      1 TAAGTCAAAAACATATATGACTTAA 25

RESULT 3
US-10-222-952A-6
; Sequence 6, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CHROML1B bottom, long oligo sequence
US-10-222-952A-6

Query Match      100.0%; Score 25; DB 15; Length 40;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TAAGTCAAAAACATATATGACTTAA 25
Db      1 TAAGTCAAAAACATATATGACTTAA 25

RESULT 4
US-10-222-952A-5/c
; Sequence 5, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20

```

```

; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 42
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CHROML1T top, long biotinylated oligo sequence; biotinylated
; OTHER INFORMATION: nucleotide at position 1
US-10-222-952A-5

Query Match      100.0%; Score 25; DB 15; Length 42;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TAAGTCAAAAACATATATGACTTAA 25
Db      42 TAAGTCAAAAACATATATGACTTAA 18

RESULT 5
US-10-282-122A-3339
; Sequence 3339, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3339
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-282-122A-3339

Query Match      100.0%; Score 25; DB 13; Length 401;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TAAGTCAAAAACATATATGACTTAA 25
|||
Db 104 TAAGTCAAAAACATATATGACTTAA 128

RESULT 6

US-10-027-632-197612
; Sequence 197612, Application US/10027632
; Publication No. US2002019837A1

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; PRIOR FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 197612

; LENGTH: 650

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-197612

Query Match 83.2%; Score 20.8; DB 13; Length 650;
Best Local Similarity 91.7%; Pred. No. 1 se+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAGTCAAAAACATATATGACTTAA 25
|||
Db 469 AACTCTAAAACATATATGACTTAA 492

RESULT 7

US-10-027-632-197612
; Sequence 197612, Application US/10027632
; Publication No. US20030204075A9

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; PRIOR FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 197612

; LENGTH: 650

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-197612

Query Match 83.2%; Score 20.8; DB 16; Length 650;

Best Local Similarity 91.7%; Pred. No. 1.6e+02;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAGTCAAAAACATATATGACTTAA 25

|||

Db 469 AACTCTAAAACATATATGACTTAA 492

RESULT 8

US-09-938-842A-3647

; Sequence 3647, Application US/09938842A

; Patent No. US20020160378A1

; GENERAL INFORMATION:

; APPLICANT: Harper, Jeff

; APPLICANT: Kreps, Joel

; APPLICANT: Wang, Xun

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

; FILE OF INVENTION: SAME, AND METHODS OF USE

; FILE REFERENCE: SRIPI300-3

; CURRENT APPLICATION NUMBER: US/09/938,842A

; PRIOR FILING DATE: 2001-08-24

; PRIOR APPLICATION NUMBER: US 60/227,866

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/264,647

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/300,111

; PRIOR FILING DATE: 2001-06-22

; NUMBER OF SEQ ID NOS: 5379

; SEQ ID NO 3647

; LENGTH: 2000

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-938-842A-3647

Query Match 80.8%; Score 20.2; DB 9; Length 2000;

Best Local Similarity 88.0%; Pred. No. 3.2e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAAGTCAAAAACATATATGACTTAA 25

|||

Db 951 TAATCATTAACATATATGACATTA 975

RESULT 9

US-09-938-842A-3647

; Sequence 3647, Application US/09938842A

; Publication No. US20040009476A9

; GENERAL INFORMATION:

; APPLICANT: Harper, Jeff

; APPLICANT: Kreps, Joel

; APPLICANT: Wang, Xun

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

; FILE OF INVENTION: SAME, AND METHODS OF USE

; FILE REFERENCE: SRIPI300-3

; CURRENT APPLICATION NUMBER: US/09/938,842A

; PRIOR FILING DATE: 2001-08-24

; PRIOR APPLICATION NUMBER: US 60/227,866

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/264,647

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/300,111

; PRIOR FILING DATE: 2001-06-22


```

; OTHER INFORMATION: Clone ID: PAT_MRT3847_93615C.1
US-10-424-599-135777

Query Match          79.2%; Score 19.8; DB 13; Length 798;
Best Local Similarity 91.3%; Pred. NO. 4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 AAGTCAAAAACATATATGACTTAA 24
        ||| ||||| ||||| ||||| |||||
DB     199 AACTCAAAAACATAGTAGCTTAA 221

RESULT 12
US-09-783-590-2768
; Sequence 2768, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillion, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2768
; LENGTH: 283
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (140)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (179)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (205)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (206)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (248)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-2768

Query Match          76.8%; Score 19.2; DB 9; Length 283;
Best Local Similarity 84.0%; Pred. NO. 5.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 TAAGTCAAAAACATATATGACTTAA 25
        ||| ||||| ||||| ||||| |||||
DB     121 TAATTCCAAAACACATATGNCITAA 145

RESULT 13
US-10-027-632-206167
; Sequence 206167, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632

```

```
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 206167
; LENGTH: 537
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(537)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-206167

Query Match
Best Local Similarity 76.8%; Score 19.2; DB 13; Length 537;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 AAGTCAAAAACATATATGACTTAA 25
    ||||| ||| ||| |||||
DB 240 AAGTCATAAAATACATGACTTAA 263

RESULT 14
US-10-027-632-206167
; Sequence 206167, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 206167
; LENGTH: 537
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(537)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-206167

Query Match
Best Local Similarity 76.8%; Score 19.2; DB 13; Length 537;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 AAGTCAAAAACATATATGACTTAA 25
    ||||| ||| ||| |||||
DB 240 AAGTCATAAAATACATGACTTAA 263

RESULT 15
US-10-424-599-106547
; Sequence 106547, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 106547
; LENGTH: 1494
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_67228C.1
US-10-424-599-106547

Query Match
Best Local Similarity 76.8%; Score 19.2; DB 13; Length 1494;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 TAAGTCAAAAACATATATGACTTA 24
    ||||| ||| ||| ||||| |||
DB 1297 TAAGTCGAAAATATATATGATTAA 1320

Search completed: May 27, 2004, 06:17:09
Job time : 516.248 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 16:18:00 ; Search time 33.4744 seconds
(without alignments)
348.146 Million cell updates/sec

Title: US-10-676-299-8
Perfect score: 21
Sequence: 1 taacacaaaagcatatgatt 21

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/prodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/prodata/2/ina/PCUS_COMB.seq.*
6: /cgn2_6/prodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.8	84.8	1026	4	US-09-134-001C-1995
2	17.8	84.8	1830121	4	US-09-557-884-1
3	17.8	84.8	1830121	4	US-09-543-990A-1
4	16.8	80.0	269223	4	US-09-596-002-41
5	16.4	78.1	678	4	US-09-543-681A-815
6	16.2	77.1	1697	1	US-08-472-028A-9
7	16.2	77.1	1697	3	US-09-071-296-9
8	16.2	77.1	1697	3	US-09-196-268-9
9	16.2	77.1	1697	3	US-09-015-683-9
10	16.2	77.1	1697	4	US-09-191-998-9
11	16.2	77.1	2235	4	US-09-328-352-3794
12	16.2	77.1	7210	4	US-09-634-238-15
13	16.2	77.1	580073	4	US-08-545-528D-1
14	16.2	77.1	1664976	4	US-08-916-421B-1
15	15.8	75.2	47	4	US-09-422-978-2280
16	15.8	75.2	202001	4	US-09-734-674-3
17	15.4	73.3	381	4	US-09-107-532A-2279
18	15.4	73.3	1386	4	US-09-543-681A-1519
19	15.4	73.3	8537	4	US-10-204-708-41
20	15.2	72.4	519	4	US-09-134-000C-724
21	15.2	72.4	627	4	US-09-669-751-43
22	15.2	72.4	678	4	US-09-328-352-2378
23	15.2	72.4	819	3	US-09-553-794-1
24	15.2	72.4	1070	4	US-09-718-841-3
25	15.2	72.4	1070	4	US-09-718-810-3
26	15.2	72.4	1070	4	US-09-718-854-3
27	15.2	72.4	1442	4	US-09-227-357-119

C 28 15.2 72.4 1689 4 US-09-718-841-1 Sequence 1, Appli
C 29 15.2 72.4 1689 4 US-09-718-810-1 Sequence 1, Appli
C 30 15.2 72.4 1689 4 US-09-718-854-1 Sequence 1, Appli
C 31 15.2 72.4 1863 4 US-09-311-826B-5 Sequence 5, Appli
C 32 15.2 72.4 2414 4 US-09-439-313-475 Sequence 475, App
C 33 15.2 72.4 2414 4 US-09-636-215-475 Sequence 475, App
C 34 15.2 72.4 2414 4 US-09-685-166A-475 Sequence 475, App
C 35 15.2 72.4 2467 1 US-08-148-581D-1 Sequence 1, Appli
C 36 15.2 72.4 5455 4 US-10-204-708-33 Sequence 33, Appli
C 37 15.2 72.4 8054 4 US-09-581-909-2 Sequence 2, Appli
C 38 15.2 72.4 319608 4 US-09-539-333D-1 Sequence 1, Appli
C 39 15.2 72.4 319608 4 US-09-679-409-1 Sequence 1, Appli
C 40 15.2 72.4 786431 4 US-09-751-389-3 Sequence 3, Appli
C 41 15.2 72.4 1664976 4 US-08-916-421B-1 Sequence 1, Appli
C 42 14.8 70.5 240 4 US-09-833-381-102 Sequence 102, Appli
C 43 14.8 70.5 522 3 US-09-342-461-5 Sequence 5, Appli
C 44 14.8 70.5 552 4 US-09-833-381-1429 Sequence 1429, Ap
C 45 14.8 70.5 630 3 US-09-328-111-408 Sequence 408, App

ALIGNMENTS

RESULT 1

US-09-134-001C-1995
; Sequence 1995, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1995
; TYPE: DNA
; LENGTH: 1026
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1995

Query Match 84.8%; Score 17.8; DB 4; Length 1026;
Best Local Similarity 90.5%; Pred. No. 9.5; Mismatches 2; Indels 0; Gaps 0;
Matches 19; Conservative 0;

Qy 1 TAACCAAAAAGCATATGATT 21
|||||
Db 602 TAACCAAAAAGCATATGATT 622

RESULT 2

US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.

TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1
Query Match 84.8%; Score 17.8; DB 4; Length 1830121;
Best Local Similarity 90.5%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TAACCAAAAGCGCATATGATT 21
DB 1238891 TAACCAAAAGCGCATATGTT 1238911
RESULT 3
US-09-643-990A-1
Sequence 1, Application US/09643990A
Patent No. 6528289
GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1
Query Match 84.8%; Score 17.8; DB 4; Length 1830121;
Best Local Similarity 90.5%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TAACCAAAAGCGCATATGATT 21
DB 1238891 TAACCAAAAGCGCATATGTT 1238911
RESULT 4
US-09-596-002-41
Sequence 41, Application US/09596002
Patent No. 6632636
GENERAL INFORMATION:
APPLICANT: Lagace, Robert, E.
APPLICANT: Patterson, Chandra
APPLICANT: Berg, Kim, L.
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
FILE REFERENCE: PM-0008-4 US
CURRENT APPLICATION NUMBER: US/09/596,002
CURRENT FILING DATE: 2000-06-16
PRIOR APPLICATION NUMBER: 60/140,121
PRIOR FILING DATE: 1999-06-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PERL Program
SEQ ID NO 41
TYPE: DNA
ORGANISM: Moraxella catarrhalis
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte template ID No. 6632636 41
PUBLICATION INFORMATION:
US-09-596-002-41
Query Match 80.0%; Score 16.8; DB 4; Length 269223;
Best Local Similarity 90.0%; Pred. No. 50;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 AACCAAAAGCGCATATGATT 21
DB 142064 AGCCAAAAGCGCATATCAIT 142083
RESULT 5
US-09-543-681A-815
Sequence 815, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRAB.
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 815
LENGTH: 678

```
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-815

Query Match
Best Local Similarity 78.1%; Score 16.4; DB 4; Length 678;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 CCAAAACGCATATGATT 21
    ||||| ||||| |||||
Db 47 CCAAAACGCATATGATT 64

RESULT 6
US-08-472-028A-9/c
; Sequence 9, Application US/08472028A
; Patent No. 5767373
; GENERAL INFORMATION:
; APPLICANT: Ward, Eric R
; APPLICANT: Volrath, Sandra
; TITLE OF INVENTION: Manipulation of Protoporphyrinogen
; TITLE OF INVENTION: Oxidase Enzyme Activity in Eukaryotic Organisms
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,028A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/261,198
; FILING DATE: 16-JUN-94
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1748/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1697 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 29..1501
; OTHER INFORMATION: /note="Arabidopsis protox-3 cDNA";
; OTHER INFORMATION: sequence from pWDC-5"
US-08-472-028A-9

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 1; Length 1697;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAACCAAAACGCATATGATT 21
    ||||| ||||| |||||
Db 1667 TAAGCTAAAGGCATATGATT 1647

RESULT 7
US-09-071-296-9/c
; Sequence 9, Application US/09196268
; Patent No. 6282837
; GENERAL INFORMATION:
; APPLICANT: Ward, Eric R
; APPLICANT: Volrath, Sandra
; TITLE OF INVENTION: Manipulation of Protoporphyrinogen
; TITLE OF INVENTION: Oxidase Enzyme Activity in Eukaryotic Organisms
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,296
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/261,198
; FILING DATE: 16-JUN-94
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1748/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1697 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 29..1501
; OTHER INFORMATION: /note="Yeast protox-3 cDNA";
; OTHER INFORMATION: sequence from pWDC-5"
US-09-071-296-9

Query Match
Best Local Similarity 85.7%; Score 16.2; DB 3; Length 1697;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAACCAAAACGCATATGATT 21
    ||||| ||||| |||||
Db 1667 TAAGCTAAAGGCATATGATT 1647

RESULT 8
US-09-196-268-9/c
; Sequence 9, Application US/09196268
; Patent No. 6282837
; GENERAL INFORMATION:
; APPLICANT: Ward, Eric R
; APPLICANT: Volrath, Sandra
; TITLE OF INVENTION: Manipulation of Protoporphyrinogen
; TITLE OF INVENTION: Oxidase Enzyme Activity in Eukaryotic Organisms
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,296
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/261,198
; FILING DATE: 16-JUN-94
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1748/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1697 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 29..1501
; OTHER INFORMATION: /note="Yeast protox-3 cDNA";
; OTHER INFORMATION: sequence from pWDC-5"
US-09-071-296-9

Query Match
Best Local Similarity 85.7%; Score 16.2; DB 3; Length 1697;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAACCAAAACGCATATGATT 21
    ||||| ||||| |||||
Db 1667 TAAGCTAAAGGCATATGATT 1647

RESULT 9
US-09-071-296-9/c
```

```
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/196.268
FILING DATE: 06-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/261,198
FILING DATE: 16-JUN-94
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1748/CIP
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1697 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 29..1501
OTHER INFORMATION: /note= "Arabidopsis protox-3 cDNA;"
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1697 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 29..1501
OTHER INFORMATION: /note= "Arabidopsis protox-3 cDNA;"
OTHER INFORMATION: sequence from pMDC-5"
US-09-015-683-9
Query Match 77.1%; Score 16.2; DB 3; Length 1697;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TAACCAAAAGGCATATGATT 21
Db 1667 TAAGCTAAAGGCATATGATT 1647
RESULT 10
US-09-191-998-9/c
Sequence 9, Application US/09191998
Patent No. 6307129
GENERAL INFORMATION:
APPLICANT: Ward, Eric R
TITLE OF INVENTION: Manipulation of Protoporphyrinogen
TITLE OF INVENTION: Oxidase Enzyme Activity in Eukaryotic Organisms
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/191,998
FILING DATE: 06-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/261,198
FILING DATE: 16-JUN-94
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1748/CIP
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1697 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 29..1501
OTHER INFORMATION: /note= "Yeast protox-3 cDNA;"
OTHER INFORMATION: sequence from pMDC-5"
US-09-196-268-9
Query Match 77.1%; Score 16.2; DB 3; Length 1697;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TAACCAAAAGGCATATGATT 21
Db 1667 TAAGCTAAAGGCATATGATT 1647
RESULT 9
US-09-015-683-9/c
Sequence 9, Application US/09015683
Patent No. 6288306
GENERAL INFORMATION:
APPLICANT: Ward, Eric R
TITLE OF INVENTION: Manipulation of Protoporphyrinogen
TITLE OF INVENTION: Oxidase Enzyme Activity in Eukaryotic Organisms
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/015,683
FILING DATE:
CLASSIFICATION:
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 29..1501
; OTHER INFORMATION: /note= "Yeast protox-3 cDNA;
; OTHER INFORMATION: sequence from pMDC-5"
US-09-191-998-9

Query Match 77.1%; Score 16.2; DB 4; Length 1697;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAACCAAAAACGCATATGATT 21
Db 1667 TAAGCTAAAGCATATGATT 1647

RESULT 11
US-09-328-352-3794
; Sequence 15, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAPMANII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3794
; LENGTH: 2235
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3794

Query Match 77.1%; Score 16.2; DB 4; Length 2235;
Best Local Similarity 85.7%; Pred. No. 61;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAACCAAAAACGCATATGATT 21
Db 1192 TTACCAAAAATGCAATTGATT 1212

RESULT 12
US-09-634-238-15
; Sequence 15, Application US/09634238
; Patent No. 6544772
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka J.
; APPLICANT: Bloksberg, Leonard, N.
; APPLICANT: Lubbers, Mark W.
; APPLICANT: Dekker, James
; APPLICANT: Christensson, Anna C.
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul W.
; APPLICANT: Reid, Julian R.
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Polynucleotides, materials incorporating
; TITLE OF INVENTION: them and methods for using them.
; FILE REFERENCE: 11000.104301
; CURRENT APPLICATION NUMBER: US/09/634,238
; CURRENT FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 7210
; TYPE: DNA
; ORGANISM: Lactobacillus rhamnosus

US-09-634-238-15
Query Match 77.1%; Score 16.2; DB 4; Length 7210;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAACCAAAAACGCATATGATT 21
Db 4301 TAACCAAAAATCGCAATGCTT 4321

RESULT 13
US-08-545-528D-1/c
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; Patent No. 6537773
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB193F1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; CURRENT FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Query Match 77.1%; Score 16.2; DB 4; Length 580073;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAACCAAAAACGCATATGATT 21
Db 159541 TAACCAAAAACGCATATGATT 159521

RESULT 14
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Built et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
; Patent No. 6503729
; TITLE OF INVENTION: Jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
```

```
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1600992)..(1600992)
OTHER INFORMATION: n equals a, t, c, or g

OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
```



```
; LOCATION: (1664854)...(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 4; Length 1664976;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAACCAAAAACGCATATGATT 21
DB 197752 TAACCAATAACGCCTATGACT 197772

RESULT 15
US-09-422-978-2280/c
; Sequence 2280, Application US/09422978
; Patent No. 6537751
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: Biallelic markers for use in constructing a high density...
; FILE REFERENCE: GENSET-020CPL
; CURRENT APPLICATION NUMBER: US/09/422,978
; CURRENT FILING DATE: 1999-10-20
; EARLIER APPLICATION NUMBER: US 09/298,850
; EARLIER FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,732
; EARLIER FILING DATE: 1998-11-23
; EARLIER APPLICATION NUMBER: US 60/082,614
; EARLIER FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 11736
; SEQ ID NO 2280
; LENGTH: 47
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 24
; OTHER INFORMATION: 99-10146-202 : polymorphic base T or A
US-09-422-978-2280

Query Match
Best Local Similarity 75.2%; Score 15.8; DB 4; Length 47;
Matches 17; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAACCAAAAACGCATATGATT 21
DB 42 TAACCAAAAATGCATATATT 22

Search completed: May 26, 2004, 17:57:01
Job time : 42.4744 secs
```

OM nucleic - nucleic search, using sw model
Run on: May 26, 2004, 16:21:09 ; Search time 1940.02 Seconds
(without alignments)
430.997 Million cell updates/sec

Title: US-10-676-299-4
Perfect score: 28
Sequence: 1 caacacataacacaaacgcatatgatt 28

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estmu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_hic.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_hic.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_estfun.*
- 16: em_estcom.*
- 17: em_gss_hum.*
- 18: em_gss_inv.*
- 19: em_gss_pin.*
- 20: em_gss_vrt.*
- 21: em_gss_fun.*
- 22: em_gss_sam.*
- 23: em_gss_mus.*
- 24: em_gss_pro.*
- 25: em_gss_rod.*
- 26: em_gss_phg.*
- 27: em_gss_vrl.*
- 28: gb_gss1.*
- 29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	23.2	82.9	677	12	BG845098 1024008G0
C 2	21.6	77.1	400	28	BH370025 AG-ND-170
C 3	21.6	77.1	814	28	BH376658 AG-ND-171
C 4	21.2	75.7	154	13	BQ704555 Bn01_02b0

C 5	21.2	75.7	497	28	AZ248907
C 6	20.6	73.6	165	28	BH193131
C 7	20.6	73.6	670	28	BH989212
8	20.6	73.6	862	29	CG919993
9	20.2	72.1	1634	10	BF966885
10	20	71.4	195	9	AV150582
C 11	20	71.4	219	12	BG586110
C 12	20	71.4	441	28	AQ18781
C 13	20	71.4	484	10	AW585883
C 14	20	71.4	524	12	BM327254
15	20	71.4	603	28	BH187898
C 16	20	71.4	603	29	CNS0751K
C 17	20	71.4	606	14	CD318965
C 18	20	71.4	611	28	AQ785876
C 19	20	71.4	631	12	BM328873
C 20	20	71.4	652	29	EX197309
C 21	20	71.4	675	29	CC625598
C 22	20	71.4	695	14	CD845514
C 23	20	71.4	701	29	AG173299
C 24	20	71.4	835	12	BG586111
C 25	20	71.4	850	28	CG249442
C 26	20	71.4	888	28	CC366342
C 27	20	71.4	889	29	CG115429
C 28	20	71.4	977	28	CC335684
C 29	20	71.4	986	29	CG115427
C 30	20	71.4	1615	11	AK043740
C 31	19.8	70.7	526	29	CG853296
C 32	19.8	70.7	545	28	BH774719
C 33	19.8	70.7	626	29	CG912896
C 34	19.8	70.7	708	29	CG411485
C 35	19.8	70.7	793	12	BI687767
C 36	19.8	70.7	837	13	BU436105
C 37	19.6	70.0	152	10	AW839294
C 38	19.6	70.0	220	10	BF813601
C 39	19.6	70.0	419	13	BU497277
C 40	19.6	70.0	423	28	BH255317
C 41	19.6	70.0	490	10	BB795977
C 42	19.6	70.0	575	29	CE075035
C 43	19.6	70.0	658	29	CE671436
C 44	19.6	70.0	666	14	CD843195
C 45	19.6	70.0	671	13	CA094832

ALIGNMENTS

RESULT 1
BG845098/c

LOCUS 1024008G02.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BG845098
VERSION BG845098.1 GI:14226282
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 677)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P., Mcdermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
TITLE Analyses of the Chlamydomonas reinhardtii genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 2
JOURNAL Unpublished (2000)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

FEATURES Location/Qualifiers

```

source
1. 677
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-1690 wild type mt+ 2igr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
II"
/notes="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP (acetate-containing) medium in the
light, TAP medium in the dark, HS (minimal) medium in
ambient levels of CO2 and HS medium bubbled with 5% CO2.
polyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with ExAssist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."

ORIGIN
Query Match 82.9%; Score 23.2; DB 12; Length 677;
Best Local Similarity 89.3%; Pred. NO. 3.4e+02;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Cy 1 CAACACATACCAACAAAGCGCATATGATT 28
|||||
Db 487 CAACACATACCAACAAAGCGCATATGACT 460
|||||

RESULT 2
BH370025/c
LOCUS
DEFINITION
AG-ND-170H24.TR ND-TAM Anopheles gambiae genomic clone
AG-ND-170H24, genomic survey sequence.
ACCESSION
BH370025
VERSION
BH370025.1 GI:17316128
KEYWORDS
GSS.
SOURCE
Anopheles gambiae (African malaria mosquito)
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE
1 (bases 1 to 400)
Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J.,
Ren,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B.,
Gardner,M.J. and Collins,F.H.
Construction of a BAC library and generation of BAC end
sequence-tagged connectors for genome sequencing of the African
malaria mosquito Anopheles gambiae
Mol. Genet. Genomics 268 (6), 720-728 (2003)
22542063
PUBMED
12655398
COMMENT
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 Rev
Class: BAC ends.
Location/Qualifiers
source
1. 814
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone_lib="ND-TAM"
/notes="Vector: pECBAC1; Site 1: HindIII"

ORIGIN
Query Match 77.1%; Score 21.6; DB 28; Length 400;
Best Local Similarity 85.7%; Pred. NO. 1.3e+03;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Cy 1 CACACATACCAACAAAGCGCATATGATT 28
|||||
Db 340 CACACATACCAACAAAGCGCATATGATT 313
|||||

RESULT 3
BH376658/c
LOCUS
DEFINITION
AG-ND-171I13.TR ND-TAM Anopheles gambiae genomic clone
AG-ND-171I13, genomic survey sequence.
ACCESSION
BH376658
VERSION
BH376658.1 GI:17322800
KEYWORDS
GSS.
SOURCE
Anopheles gambiae (African malaria mosquito)
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE
1 (bases 1 to 814)
Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J.,
Ren,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B.,
Gardner,M.J. and Collins,F.H.
Construction of a BAC library and generation of BAC end
sequence-tagged connectors for genome sequencing of the African
malaria mosquito Anopheles gambiae
Mol. Genet. Genomics 268 (6), 720-728 (2003)
22542063
PUBMED
12655398
COMMENT
Other GSSs: AG-ND-171I13.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 Rev
Class: BAC ends.
Location/Qualifiers
source
1. 814
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone_lib="ND-TAM"
/notes="Vector: pECBAC1; Site 1: HindIII"

ORIGIN
Query Match 77.1%; Score 21.6; DB 28; Length 814;

```

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 497)

AUTHORS Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., Moggan, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M. *Mouse BAC End Sequences from Library RPCI-23* Unpublished (1999)

TITLE Mouse BAC End Sequences from Library RPCI-23

JOURNAL Unpublished (1999)

COMMENT Other GSSs: RPCI-23-57D7.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 57 row: D column: 7
Seq primer: SP6
Class: BAC ends.

FEATURES

source	Location/Qualifiers
	1..497
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/strain="C57BL/6J"
	/db_xref="taxon:10090"
	/clone="RPCI-23-57D7"
	/sex="female"
	/lab_host="DH10B"
	/clone_lib="RPCI-23"
	/notes="Organ: Kidney/Brain; Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methyase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL life technologies)."

ORIGIN

Query Match	75.7%;	Score 21.2;	DB 28;	Length 497;
Best local Similarity	88.5%;	Pred. No. 1.7e+03;		
Matches	23;	Conservative	0;	Mismatches 3; Indels 0; Gaps 0;

Qy 3 ACACATAACCAAAAACGCATATGATT 28
|||||
Db 138 ACACATAACCAAGCAGCATATGATT 113
|||||

RESULT 6

LOCUS BH193131/c

DEFINITION TC3-73D19.TP TC3 Trypanosoma cruzi genomic clone TC3-73D19, genomic survey sequence.

ACCESSION BH193131

VERSION BH193131.1

KEYWORDS GI:16360848

SOURCE Trypanosoma cruzi

ORGANISM Trypanosoma cruzi

REFERENCE Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma; Schizotrypanum.

AUTHORS Kluge, S., Edwards, K.E., Nilsson, D., Bontempi, E.J., Myler, P., Stuart, K., Ghedin, E., El-Sayed, N.M. and Andersson, B.

TITLE Clustering and analysis of BAC-end and GSS sequences from Trypanosoma cruzi

COMMENT Unpublished (2001)
Other GSSs: TC3-73D19.TV

Contact: Bjorn Andersson
Department of Genetics and Pathology
Uppsala University
Rudbeck Laboratory, SE-751 85, Uppsala, Sweden
Tel: 46 18 471 4107
Fax: 46 18 471 4808
Email: bjorn.andersson@genpat.uu.se
Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
TC3. For clone availability, please contact Dr. Bjorn Andersson at
Uppsala University (bjorn.andersson@genpat.uu.se).
Seq primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers
1..165
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL Brener"
/db_xref="taxon:5693"
/clone="TC3-73D19"
/clone_lib="TC3"
/note="Vector: pBelOBAC11; Site 1: Hin dIII; Constructed
for Uppsala University by Marie-Christine Le Paslier in
the laboratory of Denis Le Paslier at the Centre d'Etude
du Polymorphisme Humain (CEPH), Paris, France. Briefly,
Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained
from Dr. Franco da Silveira) was partially digested with
Hin dIII. High molecular weight fragments were ligated in
pBelOBAC11 digested with Hin dIII. The average insert
size is 100 kb. Total clone coverage: approx. 33 X the
haploid genome."

ORIGIN

Query Match 73.6%; Score 20.6; DB 28; Length 165;
Best Local Similarity 85.2%; Pred. No. 3.3e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 CAACACATACCAAAAAGCATATGAT 27
|||||
Db 31 CAACACATACCAAAAAGCATATGAT 5

RESULT 7
BH989212 670 bp DNA linear GSS 07-OCT-2002
LOCUS
DEFINITION oeg91c01.g1 B.oleracea002 Brassica oleracea genomic, genomic survey
sequence.
ACCESSION BH989212
VERSION BH989212.1 GI:23526103
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 670)
AUTHORS Delehaunty, K., Fellwell, G., Fulton, L., McCombie, W.R., Miner, T.,
Nash, W., Rabinowicz, P.D. and Wilson, R.K.
TITLE Whole genome shotgun reads from Brassica oleracea
JOURNAL Unpublished (2002)
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submission@watson.wustl.edu
Plate: oeg91 row: c column: 01
Seq primer: -28RppOT reverse
Class: shotgun
High quality sequence start: 8
High quality sequence stop: 543.

FEATURES

Location/Qualifiers
1..670
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"

/clone_lib="B.oleracea002"
/note="Vector: pOTW13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea TO1000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."

ORIGIN

Query Match 73.6%; Score 20.6; DB 28; Length 670;
Best Local Similarity 85.2%; Pred. No. 2.5e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 2 AACACATACCAAAAAGCATATGATT 28
|||||
Db 372 AACCCATTCTCAAAAAGCATATGATT 398

RESULT 8

CG919993 862 bp DNA linear GSS 12-DRC-2003
LOCUS
DEFINITION MBED012TFC mth2 Medicago truncatula genomic clone 33B23, genomic
survey sequence.
ACCESSION CG919993
VERSION CG919993.1 GI:39779676
KEYWORDS GSS.
SOURCE Medicago truncatula (barrel medic)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE 1 (bases 1 to 862)
AUTHORS Town, C.D., Shetty, J., Koo, H. and Feldblyum, T.F.
TITLE Sequencing of BAC ends from Medicago truncatula
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: MBED012TRC
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
Seq primer: CAGGAACAGCTATGACC
Class: BAC ends.

FEATURES

Location/Qualifiers
1..862
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="33B23"
/clone_lib="mth2"
/note="Vector: pBelOBAC11; Site 1: HindIII; Site 2:
HindIII; Cook, D.R. and Kim, D.J, unpublished"

FEATURES

source
1..862
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="33B23"
/clone_lib="mth2"
/note="Vector: pBelOBAC11; Site 1: HindIII; Site 2:
HindIII; Cook, D.R. and Kim, D.J, unpublished"

ORIGIN

Query Match 73.6%; Score 20.6; DB 29; Length 862;
Best Local Similarity 85.2%; Pred. No. 2.5e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 2 AACACATACCAAAAAGCATATGATT 28
|||||
Db 647 AATACATATCCAAAAGCATATGATT 673

RESULT 9

BH966885 1634 bp mRNA linear EST 23-JAN-2001
LOCUS
DEFINITION 60228656071 NTH_MGC_95 Homo sapiens cDNA clone IMAGE:4375359 3',
mRNA sequence.
ACCESSION BH966885

```

VERSION      BP966895.1  GI:12334100
SOURCE       EST.
KEYWORDS     Homo sapiens (human)
ORGANISM     Homo sapiens
TITLE        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
COMMENT      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 1634)
AUTHORS      NIH-MGC http://mgc.ncl.nih.gov/.
TITLE        NIH-MGC National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgabs@mail.nih.gov
              Tissue procurement: Miklos Palkovits, M.D., Ph.D.
              CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
              Toshituki and Piero Carninci (RIKEN)
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM10040 Row: h column: 16
              High quality sequence start: 72
              High quality sequence stop: 252.
              Location/Qualifiers
                1..1634
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:4375359"
                  /tissue_type="hippocampus"
                  /lab_host="DH10B"
                  /clone_lib="NIH MGC 95"
                  /notes="Organ: Brain; Vector: pBluescriptR (modified
                    bluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI
                    (gtcag); Oligo-dT primed using primer
                    5'-TTTTTTTTTTTTTNN-3', size-selected for average
                    insert size 2.5 kb and normalized to R0r 5. This is a
                    primary library enriched for full-length clones and
                    constructed using the Cap-trapper method (Carninci, in
                    preparation). Library constructed by M. Brownstein
                    (NIH/NHGRI, National Institutes of Health). Note: this
                    is a NIH_MGC Library."
ORIGIN
Query Match      72.1%; Score 20.2; DB 10; Length 1634;
Best Local Similarity 88.0%; Pred. No. 3e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 CACACATACCAAAACCGATATG 25
Db 1571 CAGCACCTACCAAAACCGATATG 1595
|||||
RESULT 10
AV150582      195 bp  mRNA  linear  EST 07-JUL-1999
LOCUS         AV150582 Mus musculus hippocampus C57BL/6J adult Mus musculus cDNA
DEFINITION    Clone 2900006N19, mRNA sequence.
ACCESSION     AV150582
VERSION       AV150582.1  GI:5335788
KEYWORDS      Mus musculus (house mouse)
SOURCE        Mus musculus
ORGANISM      Mus musculus
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
              Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,
              Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T.,
              Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M.,
              Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H.,
              Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y.,
              Sugahara, Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y.,
              Tominaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T.,
              Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
              RIKEN Mouse ESTs
              Unpublished (1999)
              Contact: Chie Owa
              Genome Science Laboratory
              RIKEN
              3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
              Tel: 81-298-36-9145
              Fax: 81-298-36-9098
              Email: genome-res@rcc.riken.go.jp
              Thermostabilization and thermoactivation of thermostable enzymes by
              trehalose and its application for the synthesis of full length cDNA
              (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
              Transcriptional sequencing: A method for DNA sequencing using RNA
              polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
              Please visit our web site (http://genome.rcc.riken.go.jp) for
              further details.
              Location/Qualifiers
                1..195
                  /organism="Mus musculus"
                  /mol_type="mRNA"
                  /strain="C57BL/6J"
                  /db_xref="taxon:10090"
                  /clone="2900006N19"
                  /sex="male"
                  /tissue_type="hippocampus"
                  /dev stage="adult"
                  /clone_lib="Mus musculus hippocampus C57BL/6J adult"
ORIGIN
Query Match      71.4%; Score 20; DB 9; Length 195;
Best Local Similarity 82.1%; Pred. NO. 5.1e+03;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 CAACACATACCAAAACCGATATGATT 28
Db 1 CCACCCACCAAAACCGATATGATT 28
|||||
RESULT 11
BG586110/c     219 bp  mRNA  linear  EST 11-APR-2001
LOCUS         BG586110 MHAM Medicago truncatula/Gloms versiforme mixed EST
DEFINITION    Library cDNA clone pMHAM-41D3 5' end, mRNA sequence.
ACCESSION     BG586110
VERSION       BG586110.1  GI:13601174
KEYWORDS      Medicago truncatula/Gloms versiforme mixed EST library
SOURCE        Medicago truncatula/Gloms versiforme mixed EST library
ORGANISM      Eukaryota; mixed EST libraries.
REFERENCE     1 (bases 1 to 219)
AUTHORS       Harrison, M.J., Liu, J., Town, C.D., Van Aken, S., Utterback, T., Cho, J.
              and Fraser, C.M.
              ESTs from roots of Medicago truncatula after colonization with
              Gloms versiforme, 2001
              Unpublished (2001)
              Contact: Harrison M.J.
              Plant Biology Division
              The Samuel Roberts Noble Foundation
              2510 Sam Noble Parkway, Ardmore, OK 73401
              Tel: 580-223-5810
              Fax: 580-221-7380
              Email: mjbarrison@noble.org
              Noble EST name: N385173e TIGR sequence name: MTDC014TK More
              information is available at: http://www.medicago.org
              Seq primer: SKMD (CTA GAA CTA GCG GAT CC).
              Location/Qualifiers
                1..219
                  /organism="Medicago truncatula/Gloms versiforme mixed EST
                    library"
                  /mol_type="mRNA"
                  /cultivar="Medicago truncatula genotype Al7"
                  /db_xref="taxon:119092"

```

```

/clone="pMEAM-41D3"
/tissue type="roots colonized with Glomus versiforme"
/dev stage="Roots harvested at 10, 17, 22, 31 and 38 days
post-inoculation with Glomus versiforme. The library was
made from a mixture of RNA from each of these stages."
/lab host="E. coli strain XL0LR"
/clone lib="MHAM"
/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA from
roots harvested at 10, 17, 22, 31 and 38 days
post-inoculation with Glomus versiforme. The cDNA was
directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-zap phage using Ex-assist
helper phage and propagated in XL0LR cells."

```

ORIGIN

```

Query Match 71.4%; Score 20; DB 12; Length 219;
Best Local Similarity 82.1%; Pred. No. 5e+03;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy 1 CAACATACCAACCAACGATATGATT 28
Db 217 CATCATACCAACCAACGATATGATT 190

```

RESULT 12

```

AQ818781
LOCUS
DEFINITION
HS 5014 B2 D04 SP6E RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate-590 Col-8 Row-H, genomic survey sequence.
ACCESSION
AQ818781
VERSION
AQ818781.1 GI:5781174
KEYWORDS
GSS.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 441)
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., and
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99330589
10439784
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 590 row: H column: 8
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 441.
Location/Qualifiers
1. .441
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate:590 Col-8 Row-H"
/sex="male"
/clone_lib="RPCI-11 Human Male BAC Library"

```

FEATURES

source

```

/notes="Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"

```

ORIGIN

```

Query Match 71.4%; Score 20; DB 28; Length 441;
Best Local Similarity 82.1%; Pred. No. 4.4e+03;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy 1 CAACATACCAACCAACGATATGATT 28
Db 372 CAACATACCAACCAACGATATGATT 399

```

RESULT 13

```

AW585883/c
LOCUS
DEFINITION
EST317506 MHAM Medicago truncatula/Glomus versiforme mixed EST
library cDNA clone pMHAM-39H12, mRNA sequence.
ACCESSION
AW585883
VERSION
AW585883.1 GI:7265397
KEYWORDS
EST.
SOURCE
Medicago truncatula/Glomus versiforme mixed EST library
Eukaryota; mixed EST libraries.
REFERENCE
1 (bases 1 to 484)
Harrison, M.J., Liu, J., Peng, H., Gonzales, M., Ellis, L., Town, C.D.,
Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and Fraser, C.M.
ESTs from roots of Medicago truncatula after colonization with
Glomus versiforme
Unpublished (2000)
JOURNAL
COMMENT
Contact: Maria J. Harrison
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401, USA
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org
Noble EST name: N254811e
TIGR sequence name: MTD4H42TK
More information is available at:
'http://chryslie.tamu.edu/medicago'
Seq primer: SKmod (CTA GAA CTA gtg GAT CC).
Location/Qualifiers
1. .484
/organism="Medicago truncatula/Glomus versiforme mixed EST
library"
/mol_type="mRNA"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:119092"
/clone="pMHAM-39H12"
/tissue type="roots colonized with Glomus versiforme"
/dev stage="Roots harvested at 10, 17, 22, 31 and 38 days
post-inoculation with Glomus versiforme. The library was
made from a mixture of RNA from each of these stages."
/lab host="E. coli strain XL0LR"
/clone lib="MHAM"
/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA from
roots harvested at 10, 17, 22, 31 and 38 days
post-inoculation with Glomus versiforme. The cDNA was
directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-zap phage using Ex-assist
helper phage and propagated in XL0LR cells."

```

FEATURES

source

ORIGIN

```

Query Match 71.4%; Score 20; DB 10; Length 484;
Best Local Similarity 82.1%; Pred. No. 4.4e+03;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

Search completed: May 26, 2004, 22:26:56
Job time : 1943.02 secs

ORIGIN					
deriving from the pathogen.					
Query Match	71.4%	Score 20;	DB 12;	Length 524;	
Best Local Similarity	82.1%;	Pred. No. 4.3e+03;			
Matches 23; Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0;	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 16:21:09 ; Search time 1455.01 Seconds

(without alignments)
430.997 Million cell updates/sec

Title: US-10-676-299-8

Perfect score: 21

Sequence: 1 taacacaaacgcataatgatt 21

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

BST:*

1: em_estba.*

2: em_esthum.*

3: em_esthum.*

4: em_esthum.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_hic.*

9: gb_est1.*

10: gb_est2.*

11: gb_hic.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gsl.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18.4	87.6	774	29	CC924663 t078e02ba
C 2	18.4	87.6	1101	29	AL097067 Drosophil
C 3	17.8	84.8	400	28	BH370025 AG-ND-170
C 4	17.8	84.8	571	9	AV382514 AV382514

5	17.8	84.8	586	14	CD121148
6	17.8	84.8	694	29	CG009365
7	17.8	84.8	814	28	BH376658
8	17.8	84.8	973	28	CC419273
9	17.4	82.9	354	28	AZ045495
10	17.4	82.9	475	28	AZ045618
11	17.4	82.9	491	28	AZ045411
12	17.4	82.9	574	28	AQ579547
13	17.4	82.9	655	28	AQ579574
14	17.4	82.9	663	13	BM033357
15	17.4	82.9	682	13	BM034905
16	17	81.0	670	28	BH989212
17	17	81.0	775	29	EX134638
18	16.8	80.0	69	28	BH791889
19	16.8	80.0	112	9	AL816148
20	16.8	80.0	195	9	AV150582
21	16.8	80.0	233	10	BB214496
22	16.8	80.0	248	28	BZ766791
23	16.8	80.0	257	9	AV373538
24	16.8	80.0	257	10	BBS23534
25	16.8	80.0	275	10	BB373536
26	16.8	80.0	280	10	BB420422
27	16.8	80.0	284	28	BH559326
28	16.8	80.0	297	10	BB369070
29	16.8	80.0	303	10	BB058608
30	16.8	80.0	310	10	BB116985
31	16.8	80.0	370	9	AA960011
32	16.8	80.0	390	28	AQ844496
33	16.8	80.0	399	9	AI159538
34	16.8	80.0	409	28	AZ216537
35	16.8	80.0	439	9	AI467582
36	16.8	80.0	440	28	AQ176944
37	16.8	80.0	453	9	AI553422
38	16.8	80.0	478	10	BF449847
39	16.8	80.0	489	12	BQ046186
40	16.8	80.0	489	28	BH551845
41	16.8	80.0	498	9	AA575651
42	16.8	80.0	515	12	BM226209
43	16.8	80.0	517	10	BB757456
44	16.8	80.0	522	28	BH550313
45	16.8	80.0	526	29	CG853296

ALIGNMENTS

RESULT 1
CC924663/c
LOCUS t078e02ba.f1
DEFINITION CC924663.t078e02ba.f1 TAMB1 Bos taurus genomic clone t078e02ba, genomic survey sequence.
ACCESSION CC924663
VERSION CC924663.1 GI:33560002
KEYWORDS GSS.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
1 (bases 1 to 774)
Lin.S., Najjar, P., Adelson, D., Gill, C.A. and Roe, B.A.
Bovine BAC End Sequences from Library TAMB1
Unpublished (2003)
Contact: Bruce A. Roe
Advanced Center for Genome Technology
University of Oklahoma Department of Chemistry and Biochemistry
620 Parrington Oval, Room 208, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
Class: BAC ends
High quality sequence start: 39
High quality sequence stop: 668.

```

FEATURES
  source
    Location/Qualifiers
      1..774
        /organism="Bos taurus"
        /mol_type="genomic DNA"
        /strain="Angus bull T A M U Shoshone Y6 11519666"
        /db_xref="taxon:9913"
        /clone="t078e02ba"
        /sex="Male"
        /cell_type="Blood"
        /clone_lib="TAMBT"
        /note="Vector: pBelOBAC11; Site 1: HindIII; Site 2: HindIII; TAMBT Bovine BAC library (Male) produced by Texas A&M University, Department of Animal Science."

ORIGIN
  Query Match      87.6%; Score 18.4; DB 29; Length 774;
  Best Local Similarity 95.0%; Pred. No. 6.4e+02;
  Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  2 AACCAAAACCATATGATT 21
    |||||
Db  678 AACCAAAACCATATGATT 659

RESULT 2
LOCUS      CNS0020H/c
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
ACCESSION  BACN01003 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
VERSION     AL097067.1 GI:5608678
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
    Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 1101)

REFERENCE
  AUTHORS   Direct Submission
  TITLE     Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
  JOURNAL   Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

FEATURES
  source
    Location/Qualifiers
      1..1101
        /organism="Drosophila melanogaster"
        /mol_type="genomic DNA"
        /db_xref="taxon:7227"
        /clone="BACN01003"
        /clone_lib="DrosBAC"
        /plasmid="pBelOBAC11"
        /note="end : T7"

ORIGIN
  Query Match      87.6%; Score 18.4; DB 29; Length 1101;
  Best Local Similarity 95.0%; Pred. No. 6.4e+02;
  Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  2 AACCAAAACCATATGATT 21
    |||||
Db  299 AACCAAAACCATATGATT 280

RESULT 3

```

```

BH370025/c
LOCUS      BH370025
DEFINITION AG-ND-170H24.TR ND-TAM Anopheles gambiae genomic clone
ACCESSION  BH370025
VERSION     BH370025.1 GI:17316128
KEYWORDS    GSS.
SOURCE      Anopheles gambiae (African malaria mosquito)
  ORGANISM  Anopheles gambiae
    Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
  1 (bases 1 to 400)
  REFERENCE
    AUTHORS   Hong Y.S., Hogan J.R., Wang X., Sarkar A., Sim C., Loftus B.J., Ren C., Huff E.R., Carlile J.L., Black K., Zhang H.-B., Gardner M.J. and Collins F.H.
    TITLE     Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito Anopheles gambiae
    JOURNAL   Mol. Genet. Genomics 268 (6), 720-728 (2003)
    MEDLINE   22542063
    PUBMED    12655398
    COMMENT   Contact: Brendan J Loftus
    Department of Eukaryotic Genomics
    The Institute for Genomic Research
    9712 Medical Center Dr., Rockville, MD 20850, USA
    Tel: 301 838 0208
    Fax: 301 838 3543
    Email: b.loftus@tigr.org
    This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
    Seq primer: W13 Rev
    Class: BAC ends.
  Location/Qualifiers
    1..400
      /organism="Anopheles gambiae"
      /mol_type="genomic DNA"
      /strain="PEST"
      /db_xref="taxon:7165"
      /clone="AG-ND-170H24"
      /clone_lib="ND-TAM"
      /note="Vector: pECBAC1; Site_1: HindIII"

ORIGIN
  Query Match      84.8%; Score 17.9; DB 28; Length 400;
  Best Local Similarity 90.5%; Pred. No. 1.2e+03;
  Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy  1 TAACCAAAACGCATATGATT 21
    |||||
Db  333 TAACCAAAACGCATATGATT 313

RESULT 4
AV382514/c
LOCUS      AV382514
DEFINITION AV382514 Halocynthia roretzi Fertilized egg Halocynthia roretzi cDNA clone 001L03_5 5', mRNA sequence.
ACCESSION  AV382514
VERSION     AV382514.1 GI:6127571
KEYWORDS    EST.
SOURCE      Halocynthia roretzi
  ORGANISM  Halocynthia roretzi
    Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Stolidobranchia; Pyuridae; Halocynthia.
  1 (bases 1 to 571)
  REFERENCE

```

AUTHORS Makabe,K.W.
TITLE Halocynthia roretzi EST
JOURNAL Unpublished (1999)
COMMENT Contact: Kazuhiro W. Makabe
Department of Zoology, Graduate School of Science
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4095
Fax: 81-75-705-1113
Email: kwmakabe@sci.kyoto-u.ac.jp.

FEATURES
source
1. 571
/organism="Halocynthia roretzi"
/mol_type="mRNA"
/db_xref="taxon:7729"
/clone="001L03.5"
/dev_stage="Fertilized egg"
/clone_lib="Halocynthia roretzi Fertilized egg"

ORIGIN
Query Match 84.8%; Score 17.8; DB 9; Length 571;
Best Local Similarity 90.5%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 TAACCAAAACGCATATGATT 21
|||||
Db 138 TAACCAAAACGCATATGATT 118
|||||

RESULT 5
CD121148
LOCUS 586 bp mRNA linear EST 14-SEP-2003
DEFINITION ME1-0066T-L194-C09-U_B ME1-0066 Schistosoma mansoni CDNA clone
ACCESSION ME1-0066T-L194-C09-B, mRNA sequence.
VERSION CD121148.1 GI:34659284
KEYWORDS EST.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeida; Schistosomatidae; Schistosomatidae; Schistosoma.
1 (bases 1 to 586)
Verjovski-Almeida, S., DeMarco, R., Martins, B.A.L., Guimaraes, P.E.M.,
Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr.,
Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,
Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,
Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A.,
Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A.,
Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, I.,
Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,
Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.
Transcriptome analysis of the acelomate human parasite Schistosoma
mansoni
Nat. Genet. 35 (2), 148-157 (2003)

TITLE Nat. Genet. 35 (2), 148-157 (2003)
JOURNAL 22879926
MEDLINE
COMMENT Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjovski@usp.br
This sequence was derived from the FAPESP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL <http://bioinfo.iq.usp.br/schisto/>
Plate: ME1-0066T-L194 row: 9 column: C.
Location/Qualifiers
1. 586
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"

AUTHORS Makabe,K.W.
TITLE Halocynthia roretzi EST
JOURNAL Unpublished (1999)
COMMENT Contact: Kazuhiro W. Makabe
Department of Zoology, Graduate School of Science
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4095
Fax: 81-75-705-1113
Email: kwmakabe@sci.kyoto-u.ac.jp.

FEATURES
source
1. 571
/organism="Halocynthia roretzi"
/mol_type="mRNA"
/db_xref="taxon:7729"
/clone="001L03.5"
/dev_stage="Fertilized egg"
/clone_lib="Halocynthia roretzi Fertilized egg"

ORIGIN
Query Match 84.8%; Score 17.8; DB 9; Length 571;
Best Local Similarity 90.5%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 TAACCAAAACGCATATGATT 21
|||||
Db 138 TAACCAAAACGCATATGATT 118
|||||

RESULT 5
CD121148
LOCUS 586 bp mRNA linear EST 14-SEP-2003
DEFINITION ME1-0066T-L194-C09-U_B ME1-0066 Schistosoma mansoni CDNA clone
ACCESSION ME1-0066T-L194-C09-B, mRNA sequence.
VERSION CD121148.1 GI:34659284
KEYWORDS EST.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeida; Schistosomatidae; Schistosomatidae; Schistosoma.
1 (bases 1 to 586)
Verjovski-Almeida, S., DeMarco, R., Martins, B.A.L., Guimaraes, P.E.M.,
Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr.,
Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,
Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,
Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A.,
Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A.,
Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, I.,
Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,
Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.
Transcriptome analysis of the acelomate human parasite Schistosoma
mansoni
Nat. Genet. 35 (2), 148-157 (2003)

TITLE Nat. Genet. 35 (2), 148-157 (2003)
JOURNAL 22879926
MEDLINE
COMMENT Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjovski@usp.br
This sequence was derived from the FAPESP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL <http://bioinfo.iq.usp.br/schisto/>
Plate: ME1-0066T-L194 row: 9 column: C.
Location/Qualifiers
1. 586
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"

Unpublished (2003)
Other_GSSs: PUBQ87TB
Contact: Cathy Whitelaw
TIGR
7112 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: IF
Class: sheared ends.

Location/Qualifiers
1. 973
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTA414P05"
/note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
Cor selected genomic DNA library"

ORIGIN
Query Match 84.8%; Score 17.8; DB 28; Length 973;
Best Local Similarity 90.5%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAACCAAAAAGCCATATGATT 21
|||||
Db 230 TAACCAAAAAGCCATATGATT 210
|||||

RESULT 9
AZ045495 354 bp DNA linear GSS 14-MAR-2000
LOCUS
T234126b shotgun sub-library of BAC clone 10M16 Medicago truncatula
genomic clone 10M16-030, genomic survey sequence.

DEFINITION
ACCESSION AZ045495.1 GI:7239933
VERSION
KEYWORDS
SOURCE Medicago truncatula (barrel medic)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids 1; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE 1 (bases 1 to 354)
AUTHORS Kim,D., Baek,J., Lim,H., Peng,H., Ellis,L. and Cook,D.R.
TITLE BAC survey sequencing of Medicago truncatula (2000a)
JOURNAL Unpublished (2000)
COMMENT Contact: Cook DR
The Crop Biotechnology Center
Texas A&M University
Department of Plant Pathology and Microbiology, Rm 120 L.F.
Peterson Bldg, College Station, TX 77843-2132, USA
Tel: 409 845 8743
Fax: 409 862 4790
Email: dcook@ppserver.tamu.edu
Other name: BSC-2D-030; date: 3/1/00; Submitted to the Database of
Genome Survey Sequences (GSS) on 03/13/00; More information is
available at <http://chryslie.tamu.edu/medicago>.
Seq primer: pUC-D
Class: BAC subclone.

Location/Qualifiers
1. 354
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype Al7"
/db_xref="taxon:3880"
/clone="10M16-030"
/note="Vector: pUC18; BAC survey sequences were obtained
from sheared BAC DNA subcloned into the SmaI site of
pUC18. The template DNA for sequencing was obtained by PCR

Unpublished (2003)
Other_GSSs: PUBQ87TB
Contact: Cathy Whitelaw
TIGR
7112 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: IF
Class: sheared ends.

Location/Qualifiers
1. 973
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTA414P05"
/note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
Cor selected genomic DNA library"

ORIGIN
Query Match 84.8%; Score 17.8; DB 28; Length 973;
Best Local Similarity 90.5%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAACCAAAAAGCCATATGATT 21
|||||
Db 230 TAACCAAAAAGCCATATGATT 210
|||||

RESULT 9
AZ045495 354 bp DNA linear GSS 14-MAR-2000
LOCUS
T234126b shotgun sub-library of BAC clone 10M16 Medicago truncatula
genomic clone 10M16-030, genomic survey sequence.

DEFINITION
ACCESSION AZ045495.1 GI:7239933
VERSION
KEYWORDS
SOURCE Medicago truncatula (barrel medic)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids 1; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE 1 (bases 1 to 354)
AUTHORS Kim,D., Baek,J., Lim,H., Peng,H., Ellis,L. and Cook,D.R.
TITLE BAC survey sequencing of Medicago truncatula (2000a)
JOURNAL Unpublished (2000)
COMMENT Contact: Cook DR
The Crop Biotechnology Center
Texas A&M University
Department of Plant Pathology and Microbiology, Rm 120 L.F.
Peterson Bldg, College Station, TX 77843-2132, USA
Tel: 409 845 8743
Fax: 409 862 4790
Email: dcook@ppserver.tamu.edu
Other name: BSC-2D-030; date: 3/1/00; Submitted to the Database of
Genome Survey Sequences (GSS) on 03/13/00; More information is
available at <http://chryslie.tamu.edu/medicago>.
Seq primer: pUC-D
Class: BAC subclone.

Location/Qualifiers
1. 354
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype Al7"
/db_xref="taxon:3880"
/clone="10M16-030"
/note="Vector: pUC18; BAC survey sequences were obtained
from sheared BAC DNA subcloned into the SmaI site of
pUC18. The template DNA for sequencing was obtained by PCR

using universal primers. Sequencing reactions were primed from the PUC-C primer site (CAGGAACAGCTATGACCATGATTACGA) in the pUC18 polylinker."

ORIGIN

Query Match 82.9%; Score 17.4; DB 28; Length 354;
Best Local Similarity 94.7%; Pred. No. 1.8e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 ACCAAAACGCATATGATT 21
||||| ||||||| |||||||
Db 209 ACCAACACGCATATGATT 227

RESULT 10
AZ045618
LOCUS
DEFINITION
T234299b shotgun sub-library of BAC clone 12G15 Medicago truncatula genomic clone 12G15-011, genomic survey sequence.

ACCESSION
AZ045618
VERSION
AZ045618.1 GI:7240056
SOURCE
Medicago truncatula (barrel medic)
ORGANISM
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

REFERENCE
1 (bases 1 to 475)
AUTHORS
Kim,D., Baek,J., Lim,H., Peng,H., Ellis,L. and Cook,D.R.
TITLE
BAC survey sequencing of Medicago truncatula (2000a)
JOURNAL
Unpublished (2000)
COMMENT
Contact: Cook DR
The Crop Biotechnology Center
Texas A&M University
Department of Plant Pathology and Microbiology, Rm 120 L.F.
Peterson Bldg, College Station, TX 77843-2132, USA
Tel: 409 845 8743
Fax: 409 862 4790
Email: dcook@pserver.tamu.edu
Other name: BSC-3D-011; date: 3/1/00; Submitted to the Database of Genome Survey Sequences (GSS) on 03/13/00; More information is available at http://chryslie.tamu.edu/medicago.
Seq primer: PUC-D
Class: BAC subclone.

FEATURES

source
Location/Qualifiers
1..475
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype AL7"
/db_xref="taxon:3880"
/clone="12G15-011"
/clone_lib="shotgun sub-library of BAC clone 12G15"
/note="Vector: pUC18; BAC survey sequences were obtained from sheared BAC DNA subcloned into the SmaI site of pUC18. The template DNA for sequencing was obtained by PCR using universal primers. Sequencing reactions were primed from the PUC-C primer site (CAGGAACAGCTATGACCATGATTACGA) in the pUC18 polylinker."

ORIGIN

Query Match 82.9%; Score 17.4; DB 28; Length 475;
Best Local Similarity 94.7%; Pred. No. 1.8e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 ACCAAAACGCATATGATT 21
||||| ||||||| |||||||
Db 275 ACCAACACGCATATGATT 293

RESULT 11
AZ045411
LOCUS
AZ045411 linear GSS 14-MAR-2000

DEFINITION

ACCESSION
AZ045411
VERSION
AZ045411.1 GI:7239849
SOURCE
Medicago truncatula (barrel medic)
ORGANISM
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

REFERENCE
1 (bases 1 to 491)
AUTHORS
Kim,D., Baek,J., Lim,H., Peng,H., Ellis,L. and Cook,D.R.
TITLE
BAC survey sequencing of Medicago truncatula (2000a)
JOURNAL
Unpublished (2000)
COMMENT
Contact: Cook DR
The Crop Biotechnology Center
Texas A&M University
Department of Plant Pathology and Microbiology, Rm 120 L.F.
Peterson Bldg, College Station, TX 77843-2132, USA
Tel: 409 845 8743
Fax: 409 862 4790
Email: dcook@pserver.tamu.edu
Other name: BSC-2C-004; date: 3/1/00; Submitted to the Database of Genome Survey Sequences (GSS) on 03/13/00; More information is available at http://chryslie.tamu.edu/medicago.
Seq primer: PUC-C
Class: BAC subclone.

FEATURES

source
Location/Qualifiers
1..491
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype AL7"
/db_xref="taxon:3880"
/clone="10M16-004"
/clone_lib="shotgun sub-library of BAC clone 10M16"
/note="Vector: pUC18; BAC survey sequences were obtained from sheared BAC DNA subcloned into the SmaI site of pUC18. The template DNA for sequencing was obtained by PCR using universal primers. Sequencing reactions were primed from the PUC-C primer site (CAGGAACAGCTATGACCATGATTACGA) in the pUC18 polylinker."

ORIGIN

Query Match 82.9%; Score 17.4; DB 28; Length 491;
Best Local Similarity 94.7%; Pred. No. 1.8e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 ACCAAAACGCATATGATT 21
||||| ||||||| |||||||
Db 400 ACCAACACGCATATGATT 418

RESULT 12
AQ579547/c
LOCUS
AQ579547 linear GSS 27-SEP-1999

ACCESSION
AQ579547
VERSION
AQ579547.1 GI:4979622
SOURCE
Medicago truncatula (barrel medic)
ORGANISM
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

REFERENCE
1 (bases 1 to 574)
AUTHORS
Kim,D., Peng,H., Ellis,L. and Cook,D.R.
TITLE
BAC survey sequencing of Medicago truncatula
JOURNAL
Unpublished (1999)
COMMENT
Contact: Cook DR
The Crop Biotechnology Center

Texas A&M University
Department of Plant Pathology and Microbiology, Rm 120 L.F.
Peterson Bldg, College Station, TX 77843-2132, USA
Tel: 409 845 8743
Fax: 409 862 4790
Email: dcooke@ppserver.tamu.edu
Date: 3/3/99; Submitted to the Database of
Genome Survey Sequences (GSS) on 06/01/99; More information is
available at 'http://chrysis.tamu.edu/medicago'.
Seq primer: pUC-C
Class: BAC subclone.

FEATURES

source
Location/Qualifiers
1..574
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="10-M-16-C-008"
/clone_lib="shotgun sub-library of BAC clone 10M16"
/note="Vector: pUC18; BAC survey sequences were obtained
from sheared BAC DNA subcloned into the SmaI site of
pUC18. The template DNA for sequencing was obtained by PCR
using universal primers. Sequencing reactions were primed
from the PUC-C primer site (CAGGAACAGCTATGACCATGATTACGA)
in the pUC18 polylinker."

ORIGIN

Query Match 82.9%; Score 17.4; DB 28; Length 574;
Best Local Similarity 94.7%; Pred. No. 1.8e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ACCAAACGCTATGATT 21
|||||
Db 287 ACCAAACGCTATGATT 269
|||||

RESULT 13
LOCUS
DEFINITION
T135037b shotgun sub-library of BAC clone 10M16 Medicago truncatula
genomic clone 10-M-16-C-037, genomic survey sequence.

ACCESSION
VERSION
A0579574
A0579574.1 GI:4979649
GSS.

SOURCE
Medicago truncatula (barrel medic)

ORGANISM
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE
1 (bases 1 to 655)
Kim, D., Peng, H., Ellis, L. and Cook, D.R.
BAC survey sequencing of Medicago truncatula
Unpublished (1999)

JOURNAL
The Crop Biotechnology Center
Texas A&M University
Department of Plant Pathology and Microbiology, Rm 120 L.F.
Peterson Bldg, College Station, TX 77843-2132, USA
Tel: 409 845 8743
Fax: 409 862 4790

Email: dcooke@ppserver.tamu.edu
Other name: SSC-2-37; date: 3/3/99; Submitted to the Database of
Genome Survey Sequences (GSS) on 06/01/99; More information is
available at 'http://chrysis.tamu.edu/medicago'.
Seq primer: pUC-C
Class: BAC subclone.

FEATURES

source
Location/Qualifiers
1..655
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"

/clone="10-M-16-C-037"

/clone_lib="shotgun sub-library of BAC clone 10M16"
/note="Vector: pUC18; BAC survey sequences were obtained
from sheared BAC DNA subcloned into the SmaI site of
pUC18. The template DNA for sequencing was obtained by PCR
using universal primers. Sequencing reactions were primed
from the PUC-C primer site (CAGGAACAGCTATGACCATGATTACGA)
in the pUC18 polylinker."

ORIGIN

Query Match 82.9%; Score 17.4; DB 28; Length 655;
Best Local Similarity 94.7%; Pred. No. 1.8e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ACCAAACGCTATGATT 21
|||||
Db 367 ACCAAACGCTATGATT 385
|||||

RESULT 14
LOCUS
DEFINITION
BW035357 Nori Satoh unpublished cDNA library, blood cells Ciona
intestinalis cDNA clone cibd028110 5', mRNA sequence.

ACCESSION
VERSION
BW035357
BW035357.1 GI:23951309
EST.

KEYWORDS
SOURCE
Ciona intestinalis

ORGANISM
Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cionidae; Ciona.

REFERENCE
AUTHORS
1 (bases 1 to 663)
Satoh, Y., Satake, M., Azumi, K., Nonaka, M., Shin-i, T., Kohara, Y. and
Satoh, N.

TITLE
JOURNAL
COMMENT
Expressed genes in Ciona intestinalis (2002)
Unpublished (2002)
Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES

source
1..663
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="cibd028110"
/tissue_type="blood cells"
/clone_lib="Nori Satoh unpublished cDNA library, blood
cells"

ORIGIN

Query Match 82.9%; Score 17.4; DB 13; Length 663;
Best Local Similarity 94.7%; Pred. No. 1.8e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ACCAAACGCTATGATT 21
|||||
Db 441 ACCAAACGCTATGATT 423
|||||

RESULT 15
LOCUS
DEFINITION
BW034905 Nori Satoh unpublished cDNA library, blood cells Ciona
intestinalis cDNA clone cibd028110 5', mRNA sequence.

ACCESSION
VERSION
BW034905
BW034905.1 GI:23950857
EST.
KEYWORDS
SOURCE
Ciona intestinalis
ORGANISM
Ciona intestinalis

Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cloniidae; Ciona.

REFERENCE

1 (bases 1 to 682)

AUTHORS

Sato, Y., Satake, M., Azumi, K., Nonaka, M., Shin-i, T., Kohara, Y. and
Sato, N.

Expressed genes in Ciona intestinalis (2002)

JOURNAL

Unpublished (2002)

COMMENT

Contact: Nori Sato

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: sato@ascidian.zool.kyoto-u.ac.jp.

FEATURES

source

1..682
Location/Qualifiers
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="cibd027502"
/tissue_type="blood cells"
/clone_lib="Nori Sato unpublished cDNA library, blood
cells"

ORIGIN

Query Match

82.9%; Score 17.4; DB 13; Length 682;

Best Local Similarity

94.7%; Pred. No. 1.8e+03;

Matches

18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

3 ACCAAAAACGCATATGATT 21

|||||

Db

460 ACCAAAAACGCATATGATT 442

Search completed: May 26, 2004, 22:27:10

Job time : 1457.01 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 16:21:09 ; Search time 1732.16 Seconds
(without alignments)
430.997 Million cell updates/sec

Title: US-10-676-299-9
Perfect score: 25
Sequence: 1 ttaagtcataatgttttgactta 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estinu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pbg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20.8	83.2	580	29	CC770012 CH240_134
C 2	20.2	80.8	405	28	CC455370 SALK_0814
3	20.2	80.8	691	28	BH944102 maj94d12
C 4	20.2	80.8	904	28	BZ499748 BONGK65TR

5	19.8	79.2	370	28	BH908958
6	19.8	79.2	427	12	BM521656
7	19.8	79.2	433	10	BF596529
8	19.8	79.2	495	28	CC056262
9	19.8	79.2	666	14	CD391412
C 10	19.2	76.8	233	12	BM307747
C 11	19.2	76.8	262	9	AV538859
C 12	19.2	76.8	306	9	AV330260
C 13	19.2	76.8	371	13	BM752631
C 14	19.2	76.8	441	28	AQ986840
15	19.2	76.8	481	10	AW458707
16	19.2	76.8	483	14	CD452343
C 17	19.2	76.8	525	12	BM568294
C 18	19.2	76.8	561	13	BUI199436
C 19	19.2	76.8	609	14	CF075547
20	19.2	76.8	619	28	AZ315483
21	19.2	76.8	626	12	BI064609
C 22	19.2	76.8	627	28	AQ834629
23	19.2	76.8	634	29	AG160224
24	19.2	76.8	640	10	BE824180
C 25	19.2	76.8	657	12	BI107403
C 26	19.2	76.8	699	28	BH083564
27	19.2	76.8	700	29	CE423849
28	19.2	76.8	701	28	BZ022454
C 29	19.2	76.8	728	28	CC321273
C 30	19.2	76.8	736	28	CC310206
C 31	19.2	76.8	744	28	CC315149
C 32	19.2	76.8	763	10	BF158262
C 33	19.2	76.8	784	28	BH022212
C 34	19.2	76.8	843	28	CC080116
C 35	19.2	76.8	889	10	BF162772
C 36	19.2	76.8	926	14	CNI93829
C 37	19.2	76.8	951	29	CNS0488V
C 38	19	76.0	713	29	CE366625
C 39	19	76.0	869	28	B08279
40	18.8	75.2	420	9	AA165855
41	18.8	75.2	451	28	AQ592192
42	18.8	75.2	605	13	BU497097
43	18.8	75.2	739	12	BI328203
C 44	18.8	75.2	842	28	BZ959856
C 45	18.8	75.2	945	12	BG336309

ALIGNMENTS

RESULT 1
CC770012
LOCUS CH240_134H2.TV CHORI-240 Bos taurus Genomic Clone CH240_134H2,
DEFINITION Genomic survey sequence.
ACCESSION CC770012
VERSION CC770012.1 GI:32318930
KEYWORDS GSS.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 580)
AUTHORS Larkin,D.M., Everts-van der Wind,A., Rebeiz,M., Schweitzer,P.,
Bachman,S., Green,S., Campos,E.J., Benson,L.D., Edwards,J., Liu,L.,
Womack,J.E., de Jong,P.J. and Lewin,H.A.
Bovine BAC end sequences from CHORI-240 library
Unpublished (2003)
TITLE Other GSSs: CH240_134H2.TV
JOURNAL Contact: Harris Lewin
COMMENT Department of Animal Sciences
University of Illinois at Urbana Champaign
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 217 533 5598
Fax: 217 244 5617
Email: h-lewin@uiuc.edu

Clones are derived from the bovine BAC library CHORI-240
(http://www.chori.org/bacpac/hovine240.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering/information.html). This work
was undertaken as part of the International Bovine BAC Mapping
Consortium (IBBMC) by the University of Illinois at Urbana
Champaign, USA with funds provided by grant No. AG202-34480-11828
from USDA-CSREES and AG99-35205-8534 from USDA/NRI (Livestock
Genome Sequencing Initiative)
Plate: 134 row: H column: 2
Seq primer: SP6
Class: BAC ends.

FEATURES
source

Location/Qualifiers
1..580
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_134H2"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/note="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI;
Hereford bull U1 Domino 99375; CHORI-240 Bovine BAC
Library (Male) produced by Pieter de Jong"

ORIGIN

Query Match 83.2%; Score 20.8; DB 29; Length 580;
Best Local Similarity 91.7%; Pred. No. 1.1e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTGGACTT 24
||||| ||||| ||||| ||||| |||||
DB 555 TTAAGTCAATATGTTTGGACTT 578

RESULT 2
CC455370/c

LOCUS
DEFINITION
SALK_081478.48.30.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_081478.48.30.x, genomic
survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CC455370 405 bp DNA linear GSS 30-MAY-2003
SALK_081478.48.30.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_081478.48.30.x, genomic
survey sequence.
CC455370.1 GI:31215393
GSS
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 405)

REFERENCE
AUTHORS

Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
Gadranab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
Shinn,P., Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Unpublished (2001)

TITLE

Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.

Class: TDNA tagged.

Location/Qualifiers

1..405
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"

FEATURES
source

/db_xref="taxon:3702"
/clone="SALK_081478.48.30.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN

Query Match 80.8%; Score 20.2; DB 28; Length 405;
Best Local Similarity 88.0%; Pred. No. 1.9e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTGGACTTA 25
||||| ||||| ||||| ||||| |||||
DB 168 TTAAGTCATATATGTTTGGACTTA 144

RESULT 3

BH944102

LOCUS

DEFINITION
BH944102.g1 B.oleracea002 Brassica oleracea genomic, genomic survey
sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BH944102 691 bp DNA linear GSS 01-OCT-2002
BH944102.g1 B.oleracea002 Brassica oleracea genomic, genomic survey
sequence.
BH944102.1 GI:23424162
GSS.
Brassica oleracea
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 691)
Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,
Nash,W., Rabinowicz,P.D. and Wilson,R.K.
Whole genome shotgun reads from Brassica oleracea
Unpublished (2002)
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Plate: maj94 row: d column: 12
Seq primer: -28Rppot reverse
Class: shotgun
High quality sequence start: 56
High quality sequence stop: 535.
Location/Qualifiers
1..691
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTW13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea T01000DH3 buds provided by
Thomas Osborn at the University of Wisconsin Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."

FEATURES
source

1..691
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTW13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea T01000DH3 buds provided by
Thomas Osborn at the University of Wisconsin Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."

ORIGIN

Query Match 80.8%; Score 20.2; DB 28; Length 691;
Best Local Similarity 88.0%; Pred. No. 1.7e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTGGACTTA 25
||||| ||||| ||||| ||||| |||||
DB 453 TTAAGTCATATATGTTTGGACTTA 477

RESULT 4

BZ499748/c

LOCUS BZ499748 904 bp DNA linear GSS 16-DEC-2002
DEFINITION BONGK65TR BO_1.6_2_KB tot Brassica oleracea genomic clone BONGK65,
 genomic survey sequence.
ACCESSION BZ499748
VERSION BZ499748.1 GI:27015132
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 904)
AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other GSSs: BONGK65TF
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.
FEATURES
 source
 1..904
 Location/Qualifiers
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO100DH3"
 /db_xref="taxon:3712"
 /clone="BONGK65"
 /clone_lib="BO_1.6_2_KB tot"
 /note="vector: PHOS1; Site 1: BstXI; 1.6-2 kb sheared
 total DNA inserted into PHOS1 using BstXI linkers"
ORIGIN
 Query Match 80.8%; Score 20.2; DB 28; Length 904;
 Best Local Similarity 88.0%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 TTAAGTCATATATGTTTGACTTA 25
 ||||| ||||| ||||| ||||| |||||
 DB 292 TTAAGTTAAATATGTTTGACTTA 268
RESULT 5
 BH908958
LOCUS BH908958 370 bp DNA linear GSS 04-SEP-2002
DEFINITION SALK 051535.31.15.x Arabidopsis thaliana TDNA insertion lines
 Arabidopsis thaliana genomic clone SALK_051535.31.15.x, genomic
 survey sequence.
ACCESSION BH908958
VERSION BH908958.1 GI:22721891
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 370)
AUTHORS Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,
 Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,
 Shinn, P., Zimmerman, J. and Ecker, J.R.
TITLE A Sequence-Indexed Library of Insertion Mutations in the
 Arabidopsis Genome
JOURNAL Unpublished (2001)
COMMENT Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379

Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of
 TDNA.
 Class: TDNA tagged.
FEATURES
 source
 1..370
 Location/Qualifiers
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_051535.31.15.x"
 /clone_lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines
 each of which contains one or more TDNA insertion
 elements. The resultant fragment for each line was
 directly sequenced to determine the genomic sequence at
 the site of insertion. Details of the protocols used can
 be found at http://signal.salk.edu/tdna_protocols.html"
ORIGIN
 Query Match 79.2%; Score 19.8; DB 28; Length 370;
 Best Local Similarity 91.3%; Pred. No. 2.6e+03;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 TTAAGTCATATATGTTTGACT 23
 ||||| ||||| ||||| ||||| |||||
 DB 4 TTAATCATATATATTTTGACT 26
RESULT 6
 BM521656
LOCUS BM521656 427 bp mRNA linear EST 15-FEB-2002
DEFINITION sak60404.Y1 Gm-cl036 Glycine max cDNA clone SOYBEAN CLONE ID:
 Gm-cl036-7544 5' similar to TR:Q9SQUT Q9SQUT P24P17.7 PROTEIN. ;
 mRNA sequence.
ACCESSION BM521656
VERSION BM521656.1 GI:18692808
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max (soybean)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
REFERENCE 1 (bases 1 to 427)
AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V.,
 Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,
 Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,
 Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N.,
 Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
 McCann, R., Waterston, R. and Wilson, R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: ResGen, Invitrogen Corp. 2130
 South Memorial Parkway Hunttsville, AL 35801 For further information
 call: (800)-533-4363 or contact: ccu@resgen.com web site:
 www.resgen.com
 Putative full length read
 vector to vector length is 543
 Seq primer: -40RP from Gibco.
FEATURES
 source
 1..427
 Location/Qualifiers
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="SOYBEAN CLONE ID: Gm-cl036-7544"

```

/tissue_type="somatic embryos cultured on MSD 20"
/lab_host="DH10B"
/clone_lib="Gm-cl036"
/notes="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; This
cDNA library was constructed from mRNA isolated from
somatic embryos (age ranging from 2 months to 9 months)
cultured on MSD 20. The library was prepared using the
Life Technologies pSuperScript cDNA library construction
kit. Complementary DNA was synthesized from mRNA using a
poly (dT) sequence with a NotI restriction site. SalI
linkers adapters were ligated to the blunt-ended cDNA
fragments followed by NotI digestion. The cDNA fragments
were directionally cloned into the NotI-SalI restriction
site of the pSPORT1 vector. The ligated cDNA fragments
were transformed into E.coli ElectroMax DH10B host cells.
This library was constructed in the laboratory of Dr. Lila
Vodkin by Anu Khanna at the University of Illinois at
Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

```

ORIGIN

```

Query Match      79.2%; Score 19.8; DB 12; Length 427;
Best Local Similarity 91.3%; Pred. No. 2.6e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy  2 TAAGTCATATATGTTTGGACTT 24
    |||||
Db  388 TAAGTCATATGTTTGGACTT 410

```

```

RESULT 7
BF596529
LOCUS
DEFINITION
su72e06.y1 Gm-cl055 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl055-227.5, similar to TR:Q95QU7 Q95QU7 F24P17.7 PROTEIN. ;,
mRNA sequence.

```

```

ACCESSION
BF596529
VERSION
BF596529.1 GI:11688853
KEYWORDS
EST.
SOURCE
Glycine max (soybean)
ORGANISM
Glycine max

```

```

REFERENCE
1 (bases 1 to 433)
Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V.,
Khanna,A., Bolls,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
McCann,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Putative full length read
Vector to vector length is 533 This clone is available through:
ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL
35801 For further information call: (800)-533-4363 or contact via
email: ccr@resgen.com
High quality sequence stop: 424.
Location/Qualifiers
1..433
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl055-227"
/tissue_type="Mature seed pods, Greenhouse grown"

```

FEATURES

```

source
Location/Qualifiers
1..433
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl055-227"
/tissue_type="Mature seed pods, Greenhouse grown"

```

```

/lab_host="DH10B"
/clone_lib="Gm-cl055"
/notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from mRNA isolated
from mature seed pods of greenhouse grown plants prior to
senescence for the cultivar KPI. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with a XhoI restriction site. EcoRI
adapters were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA fragments were
directionally cloned into the EcoRI-XhoI restriction site
of the pBluescript vector. The ligated cDNA fragments
were transformed into DH10B host cells (GibcoBRL). This
library was constructed in the laboratory of Dr. Randy
Shoemaker."

```

ORIGIN

```

Query Match      79.2%; Score 19.8; DB 10; Length 433;
Best Local Similarity 91.3%; Pred. No. 2.6e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy  2 TAAGTCATATATGTTTGGACTT 24
    |||||
Db  369 TAAGTCATATGTTTGGACTT 391

```

```

RESULT 8
CC056262
LOCUS
DEFINITION
SALK_101695.17.45.x Arabidopsis thaliana genomic clone SALK_101695.17.45.x, genomic
survey sequence.

```

```

ACCESSION
CC056262
VERSION
CC056262.1 GI:29475926
KEYWORDS
GSS.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana

```

```

REFERENCE
1 (bases 1 to 495)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
Gadranab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
Shinn,P., Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA. This sequence lies within an annotated intron of At3g28260.
Class: TDNA tagged.
Location/Qualifiers
1..495
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_101695.17.45.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/notes="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna\_protocols.html"

```

REFERENCE

```

AUTHORS
Shinn,P., Zimmerman,J. and Ecker,J.R.
TITLE
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
JOURNAL
Unpublished (2001)
COMMENT
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA. This sequence lies within an annotated intron of At3g28260.
Class: TDNA tagged.
Location/Qualifiers
1..495
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_101695.17.45.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/notes="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna\_protocols.html"

```

FEATURES

```

source
Location/Qualifiers
1..495
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_101695.17.45.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/notes="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna\_protocols.html"

```

ORIGIN

```

Query Match      79.2%; Score 19.8; DB 28; Length 495;
Best Local Similarity 91.3%; Pred. No. 2.5e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAGTCATATATGTTTGGACTTA 25
    |||||
Db 202 AAGTAATATATGTTTGGACTTA 224

RESULT 9
LOCUS CD391412/c
DEFINITION Gm ck10293 Soybean induced by Salicylic Acid Glycine max cDNA 3',
mRNA sequence.
ACCESSION CD391412
VERSION CD391412.1 GI:31306209
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE 1 (bases 1 to 666)
AUTHORS Tian,A.-G., Wang,J., Cui,P., Han,Y.-J., Xu,H., Cong,L.-J.,
Huang,X.-G., Wang,X.-L., Jiao,Y.-Z., Wang,B.-J., Wang,Y.-J.,
Zhang,J.-S., Chen,S.-Y. and Yu,J.
TITLE Soybean Expressed Sequence Tags Sequencing
JOURNAL Unpublished (2003)
COMMENT Contact: Chen S-Y
Plant Biotechnology Laboratory
Institute of Genetics and Developmental Biology, CAS, China
Datun road, Beijing 100101, China
Tel: 86-10-64886859
Fax: 86-10-64873428
Email: sychen@genetics.ac.cn
Email: sychen@genetics.ac.cn
Seq primer: 17 primer.
FEATURES
source
1..666
Location/Qualifiers
/organism="Glycine max"
/mol_type="mRNA"
/cultivar="Kefeng 1"
/db_xref="taxon:3847"
/tissue_type="Seedlings"
/dev_stage="two-week seedlings"
/clone_lib="XLI-Blue MRF, strain"
/clone_lib="Soybean induced by Salicylic Acid"
/notes="Vector: pBluescript SK+; Site 1: EcoR I; Site 2:
Xho I. The cDNA library was constructed by He, C-Y from
mRNA isolated from two-week seedlings (cultivar Kefeng 1)
treated by spraying 2.0mM salicylic acid for 24, 36, 48
and 72 h. Complementary DNA was synthesized from mRNA
using a primer consisting of a poly(dT) sequence with a
XhoI restriction site. EcoRI adapters were ligated to the
blunt-ended cDNA fragments followed by XhoI digestion. The
cDNA fragments were directionally cloned into the
EcoRI-XhoI restriction site of the pBluescript vector. The
ligated cDNA fragments were transformed into XLI-Blue MRF
host cells (Stratagene)."
```

```

Query Match      79.2%; Score 19.8; DB 14; Length 666;
Best Local Similarity 91.3%; Pred. No. 2.4e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TAAGTCATATATGTTTGGACTT 24
    |||||
Db 315 TAAGTCATATATGTTTGGACTT 293

RESULT 10
LOCUS BM307747/c
DEFINITION BM307747
ACCESSION BM307747
VERSION BM307747.1 GI:18039453
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE 1 (bases 1 to 233)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Expelding,J., Corvett,V.,
Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
Bowers,Y., Person,B., Swaller,F., Gibbons,M., Pape,D., Harvey,N.,
Schurk,R., Ritter,B., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact: ccu@resgen.com web site:
www.resgen.com
Seq primer: -40RP from Gibco
High quality sequence stop: 223.
FEATURES
source
1..233
Location/Qualifiers
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl075-4777"
/tissue_type="differentiating somatic embryos cultured on
MSM6AC"
/lab_host="DH10B"
/clone_lib="Gm-cl075"
/notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from mRNA isolated
from differentiating somatic embryos cultured on MSM6AC.
The library was prepared using the Stratagene pBluescript
II SK(+) library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with an XhoI restriction site. EcoRI
adaptors were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA fragments were
directionally cloned into the EcoRI-XhoI restriction site
of the pBluescript vector. The ligated cDNA fragments
were transformed into E.coli Electromax DH10B host cells.
Tissue culture and library construction were performed by
Francoise Thibaud-Nissen and Anu Khana (Lila Vodkin lab,
University of Illinois)."
```

```

Query Match      76.8%; Score 19.2; DB 12; Length 233;
Best Local Similarity 87.5%; Pred. No. 4.6e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TAAGTCATATATGTTTGGACTTA 25
    |||||
Db 59 TAAATCATATATATTTTCGACTTA 36

RESULT 11
LOCUS AV538859
DEFINITION AV538859/c
DEFINITION AV538859
ACCESSION AV538859
VERSION AV538859.1 GI:18039453
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE 1 (bases 1 to 262)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Expelding,J., Corvett,V.,
Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
Bowers,Y., Person,B., Swaller,F., Gibbons,M., Pape,D., Harvey,N.,
Schurk,R., Ritter,B., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact: ccu@resgen.com web site:
www.resgen.com
Seq primer: -40RP from Gibco
High quality sequence stop: 223.
FEATURES
source
1..262
Location/Qualifiers
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl075-4777"
/tissue_type="differentiating somatic embryos cultured on
MSM6AC"
/lab_host="DH10B"
/clone_lib="Gm-cl075"
/notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from mRNA isolated
from differentiating somatic embryos cultured on MSM6AC.
The library was prepared using the Stratagene pBluescript
II SK(+) library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with an XhoI restriction site. EcoRI
adaptors were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA fragments were
directionally cloned into the EcoRI-XhoI restriction site
of the pBluescript vector. The ligated cDNA fragments
were transformed into E.coli Electromax DH10B host cells.
Tissue culture and library construction were performed by
Francoise Thibaud-Nissen and Anu Khana (Lila Vodkin lab,
University of Illinois)."
```

The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Sushiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,
Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and
Hayashizaki, Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 95 (7), 3455-3460 (1998)
Itoh, M., Kikunai, T., Akiyama, K., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,
Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. *Genome Res.* 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp/>) for
further details.

```
Location/Qualifiers
1. 306
/organism="Mus musculus"
/mol type="mRNA"
```

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="6330512K13"
/sex="male"
/tissue_type="medulla oblongata"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult male medulla oblongata"

```

/note="Site 1: Sall; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse

transcriptase and subsequently enriched for full-length by a cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGACAGAGATTCGAGTATTAATAATTCCTCCGCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

```

  76.8%; Score 19.2; DB 9; Length 306;
  Similarity 87.5%; Pred. NO. 4..4e+03;
  21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

  1 TTAAGTCATATATGTTTTCGACTT 24
    ||||| ||||| |||||
  18 TTAATTCATATATCTTCTTGACTT 25

  BX752631 371 bp mRNA linear EST 03-DEC-2003
  BX752631 XGC-gastrula silurana tropicalis cDNA clone TCas080p20 3',
  mRNA sequence.
  BX752631
  BX752631.1 GI:39642699
  EST.
  silurana tropicalis (western clawed frog)
  silurana tropicalis

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Silurana.

1 (bases 1 to 371)

Croneing, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J. Sanger Xencopus tropicalis EST project 2001 (11_2003)

Unpublished (2003)

Contact: Croneing MDR

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: tropesanger.ac.uk

Sanger Xencopus tropicalis EST project 2001

TROPICALIS SEQUENCE ID: Tgas080p20.q1k77

Sequencing primer: T7

This sequence is from a Xencopus Gene Collection (XGC) library constructed by Aaron M. Zorn.

cDNA was oligo dT primed from Sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end.

Vector: pCS107; Site 1: EcoRI; Site 2: NotI

Host: Escherichia coli XL1-blue.

FEATURES

Location/Qualifiers

1..371

/organism="Silurana tropicalis"

/mol_type="mRNA"

/db_xref="taxon:8364"

/clone="Tgas080p20"

/dev_stages="gastrula (stages 10.5-12 mixed)"

/lab_host="Escherichia coli XL1-blue"

/clone_lib="XGC-gastrula"

/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dT primed from Sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

source

ORIGIN

Query Match 76.8%; Score 19.2; DB 13; Length 371;

Best Local Similarity 87.5%; Pred. No. 4.3e+03;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTTCAGTT 24

|||||

DB 158 TTAAGTCATATATGTTTTCAGTT 135

RESULT 14

AQ986840

LOCUS

AQ986840.1 G1:6820045

DEFINITION

RPCEI-23-308D20-TV RPCEI-23 Mus musculus genomic clone

ACCESSION

AQ986840

VERSION

AQ986840.1 G1:6820045

KEYWORDS

GSS.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 441)

Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.

Mouse BAC End Sequences from Library RPCEI-23

Unpublished (1999)

Other GSSs: RPCEI-23-308D20.TU

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@igrr.org

Clones are derived from the mouse BAC library RPCEI-23. For BAC

FEATURES

Location/Qualifiers

1..441

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCEI-23-308D20"

/sex="Female"

/lab_host="DH10B"

/clone_lib="RPCEI-23"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL life Technologies)."

library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 308 row: D column: 20

Seq primer: T7

Class: BAC ends.

FEATURES

Location/Qualifiers

1..441

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCEI-23-308D20"

/sex="Female"

/lab_host="DH10B"

/clone_lib="RPCEI-23"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL life Technologies)."

ORIGIN

Query Match 76.8%; Score 19.2; DB 28; Length 441;

Best Local Similarity 87.5%; Pred. No. 4.1e+03;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTTCAGTT 24

|||||

DB 298 TTAAGTCATATATGTTTTCAGTT 321

ORIGIN

RESULT 15

AW458707

LOCUS

AW458707.1 G1:7028924

DEFINITION

SA12F03.y1 Gm-cl016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl016-4566 5', mRNA sequence.

ACCESSION

AW458707

VERSION

AW458707.1 G1:7028924

KEYWORDS

EST.

SOURCE

Glycine max (soybean)

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT

PROJECT

INSTITUTION

ADDRESS

TELEPHONE

FAX

EMAIL

WWW

OTHER

NOTES

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

```
source      1. .481
            /organism="Glycine max"
            /mol_type="mRNA"
            /db_xref="taxon:3847"
            /clone="GENOME SYSTEMS CLONE ID: Gm-cl016-4566"
            /tissue_type="immature flowers of field grown plants"
            /lab_host="XL10-Gold"
            /clone_lib="Gm-cl016"
            /note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2:
            XhoI; This cDNA library was constructed from mRNA isolated
            from immature flowers of field grown plants. The cDNA
            library was prepared using the Stratagene pBluescript II
            XR library construction kit. Complementary DNA was
            synthesized from mRNA using a primer consisting of a poly
            (dT) sequence with a XhoI restriction site. EcoRI adapters
            were ligated to the blunt-ended cDNA fragments followed by
            XhoI digestion. The cDNA fragments were directionally
            cloned into the EcoRI-XhoI restriction site of the
            pBluescript vector. The ligated cDNA fragments were
            transformed into XL10-Gold host cells. This library was
            constructed by Dr. Randy Shoemaker and Dr. John
            Erpeliding."
```

ORIGIN

```
Query Match      76.8%; Score 19.2; DB 10; Length 481;
Best Local Similarity 87.5%; Pred. No. 4.1e+03;
Matches 21; Conserved 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TAAGTCATATATGTTTGGACTTA 25
   |||||
Db 27 TAAATCATATATATTTTCGACTTA 50
```

Search completed: May 26, 2004, 22:27:15
Job time : 1737.16 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 16:21:09 ; Search time 1732.16 Seconds
(without alignments)
430.997 Million cell updates/sec

Title: US-10-676-299-10

Perfect score: 25

Sequence: 1 taagcacaataatgatgactaa 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_hct.*

9: gb_est1.*

10: gb_est2.*

11: gb_hct.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	20.8	83.2	580	29	CC770012 CH240.134
C 2	20.2	80.8	405	28	CC455370 SALK_0814
C 3	20.2	80.8	691	28	BH944102 maj94d12.
C 4	20.2	80.8	904	28	BZ499748 BONGK65FR

C 5	19.8	79.2	370	28	BH908958
C 6	19.8	79.2	427	12	BM521656
C 7	19.8	79.2	433	10	BF596529
C 8	19.8	79.2	495	28	CC056262
C 9	19.8	79.2	666	14	CD331412
10	19.2	76.8	233	12	EM307747
11	19.2	76.8	262	9	AV538859
12	19.2	76.8	306	9	AV330260
13	19.2	76.8	371	13	BF752631
C 14	19.2	76.8	441	28	AQ986840
C 15	19.2	76.8	481	10	AW458707
C 16	19.2	76.8	483	14	CD452343
C 17	19.2	76.8	525	12	EM568294
C 18	19.2	76.8	561	13	BU199436
C 19	19.2	76.8	609	14	CF075547
C 20	19.2	76.8	619	28	AZ315483
C 21	19.2	76.8	626	12	BI064609
C 22	19.2	76.8	627	28	AQ834629
C 23	19.2	76.8	634	29	AG160224
C 24	19.2	76.8	640	10	BB824180
C 25	19.2	76.8	657	12	BI107403
C 26	19.2	76.8	699	28	BH083564
C 27	19.2	76.8	700	29	CB423849
C 28	19.2	76.8	701	28	BZ022454
C 29	19.2	76.8	728	28	CC321273
C 30	19.2	76.8	736	28	CC310206
C 31	19.2	76.8	744	28	CC315149
C 32	19.2	76.8	763	10	BF168262
C 33	19.2	76.8	784	28	BH022212
C 34	19.2	76.8	843	28	CC080116
C 35	19.2	76.8	889	10	BF162772
C 36	19.2	76.8	926	14	CK193829
C 37	19.2	76.8	951	29	CNS0488V
C 38	19.2	76.8	969	28	CH366625
C 39	19.2	76.8	869	28	B08279
C 40	18.8	75.2	420	9	AA165855
C 41	18.8	75.2	451	28	AQ592192
C 42	18.8	75.2	605	13	BU497097
C 43	18.8	75.2	739	12	BI328203
C 44	18.8	75.2	842	28	BZ959856
C 45	18.8	75.2	945	12	BG336309

ALIGNMENTS

RESULT 1

CC770012/c

LOCUS

DEFINITION

Genomic survey sequence.

ACCESSION

CC770012

VERSION

CC770012.1

KEYWORDS

SOURCE

ORGANISM

Bos taurus (cow)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CC770012 580 bp DNA linear GSS 27-JUN-2003

CH240.134H2.TJ CHORI-240 Bos taurus genomic clone CH240.134H2,

Genomic survey sequence.

CC770012 GI:32318930

GSS.

Bos taurus (cow)

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 580)

Larkin, D.M., Everts-van der Wind, A., Rebeiz, M., Schweitzer, P.,

Sachman, S., Green, S., Campos, E.J., Benson, L.D., Edwards, J., Liu, L.,

McMack, J.E., de Jong, P.J. and Lewin, H.A.

Bovine BAC end sequences from CHORI-240 library

Unpublished (2003)

Other GSSs: CH240.134H2.TV

Contact: Harris Lewin

Department of Animal Sciences

University of Illinois at Urbana Champaign

1201 W. Gregory Dr., Urbana, IL 61801, USA

Tel: 217 333 5998

Fax: 217 244 5617

Email: h-lewin@uiuc.edu

Clones are derived from the bovine BAC library CHORI-240 (http://www.chori.org/bacpac/bovine240.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.choi.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/orderinginformation.htm). This work was undertaken as part of the International Bovine BAC Mapping Consortium (IBBMC) by University of Illinois at Urbana Champaign, USA with funds provided by grant No. AG202-34480-11828 from USDA-CSREES and AG99-35205-8534 from USDA/NRI (Livestock Genome Sequencing Initiative).
Plate: 134 row: H column: 2
Seq primer: SP6
Class: BAC ends.

FEATURES source

Location/Qualifiers
1..580
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_134H2"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI; Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC library (Male) produced by Pieter de Jong"

ORIGIN

Query Match 83.2%; Score 20.8; DB 29; Length 580;
Best Local Similarity 91.7%; Pred. No. 1.1e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAGTCAAAAACATATATGACTTAA 25
|||||
DB 578 AAGTCAAAAACATATATGACTTAA 555

RESULT 2

LOCUS CC455370 405 bp DNA linear GSS 30-MAY-2003
DEFINITION SALK_081478.48.30.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_081478.48.30.x, genomic survey sequence.
CC455370
CC455370.1 GI:31215393
GSS.
Arabidopsis thaliana (thale cress)

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 405)

REFERENCE

AUTHORS

Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
Shinn,P., Zimmerman,J. and Ecker,J.R.

TITLE

JOURNAL

COMMENT

A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
The Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: eckers@salk.edu
This is single pass sequence recovered from the left border of TDNA.
Class: TDNA tagged.

FEATURES

source

Location/Qualifiers
1..405
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"

/db_xref="taxon:3702"
/clone="SALK_081478.48.30.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/notes="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN

Query Match 80.8%; Score 20.2; DB 28; Length 405;
Best Local Similarity 88.0%; Pred. No. 1.9e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAAGTCAAAAACATATATGACTTAA 25
|||||
DB 144 TAAATCAATAAACATATATGACATAA 168

RESULT 3

BH944102/c

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

LOCUS BZ499748 904 bp DNA linear GSS 16-DEC-2002
 DEFINITION BONGK65TR BO_1.6_2_KB_tot Brassica oleracea genomic clone BONGK65,
 genomic survey sequence.
 ACCESSION BZ499748
 VERSION BZ499748.1 GI:27015132
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 904)
 REFERENCE Town, C.D., Van Aken, S., Utterback, T., Koo, H., and Fraser, C.M.
 AUTHORS Whole genome shotgun sequencing of Brassica oleracea
 TITLE Unpublished (2001)
 JOURNAL Other GSSs: BONGK65TF
 COMMENT Contact: Chris Town
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdow@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.
 Location/Qualifiers
 1..904
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="T0100DH3"
 /db_xref="taxon:3712"
 /clone="BONGK65"
 /clone_lib="BO_1.6_2_KB_tot"
 /note="Vector: pHOS1; Site 1: BstXI; 1.6-2 kb sheared
 total DNA inserted into pHOS1 using BstXI linkers"

Query Match 80.8%; Score 20.2; DB 28; Length 904;
 Best Local Similarity 88.0%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAAGTCAGAAACATATATGACTTAA 25
 |||||
 DB 268 TAAGTCAGAAACATATATGACTTAA 292
 |||||

RESULT 5
 BH908958/c
 LOCUS BH908958 370 bp DNA linear GSS 04-SEP-2002
 DEFINITION SALK_051535.31.15.x Arabidopsis thaliana TDNA insertion lines
 Arabidopsis thaliana genomic clone SALK_051535.31.15.x, genomic
 survey sequence.
 ACCESSION BH908958
 VERSION BH908958.1 GI:22721891
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 370)
 REFERENCE Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,
 AUTHORS Gadgil, J.C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,
 Shinn, P., Zimmerman, J., and Ecker, J.R.
 TITLE A Sequence-Indexed Library of Insertion Mutations in the
 Arabidopsis Genome
 JOURNAL Unpublished (2001)
 COMMENT Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 5379

Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of
 TDNA.
 Class: TDNA tagged.
 Location/Qualifiers
 1..370
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_051535.31.15.x"
 /clone_lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines
 each of which contains one or more TDNA insertion
 elements. The resultant fragment for each line was
 directly sequenced to determine the genomic sequence at
 the site of insertion. Details of the protocols used can
 be found at http://signal.salk.edu/tdna_protocols.html"

Query Match 79.2%; Score 19.8; DB 28; Length 370;
 Best Local Similarity 91.3%; Pred. No. 2.6e+03;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGTCAGAAACATATATGACTTAA 25
 |||||
 DB 26 AGTCAGAAACATATATGACTTAA 4
 |||||

RESULT 6
 BM521656/c
 LOCUS BM521656 427 bp mRNA linear EST 15-FEB-2002
 DEFINITION salk60404.y1 Gm-cl036 Glycine max cDNA clone SOYBEAN CLONE ID:
 Gm-cl036-7544 5' similar to TR:Q9SQ07 Q9SQ07 F24P17.7 PROTEIN. ;,
 mRNA sequence.
 ACCESSION BM521656
 VERSION BM521656.1 GI:18692808
 KEYWORDS EST.
 SOURCE Glycine max (soybean)
 ORGANISM Glycine max (soybean)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 427)
 REFERENCE Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V.,
 AUTHORS Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,
 Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,
 Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N.,
 Schurr, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
 McCann, R., Waterston, R., and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available through: ResGen, Invitrogen Corp. 2130
 South Memorial Parkway Huntsville, AL 35801 For further information
 call: (800)-533-4363 or contact: ccu@resgen.com web site:
www.resgen.com
 Putative full length read
 vector to vector length is 543
 Seq primer: -40RP from Gibco.
 Location/Qualifiers
 1..427
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="SOYBEAN CLONE ID: Gm-cl036-7544"

FEATURES

source

```
/tissue_type="somatic embryos cultured on MSD 20"
/lab_host="DH10B"
/clone_lib="Gm-cl036"
/notes="Vector: pSPOR1; Site 1: NotI; Site 2: SalI; This
cDNA library was constructed from mRNA isolated from
somatic embryos (age ranging from 2 months to 9 months)
cultured on MSD 20. The library was prepared using the
Life Technologies psuperScript cDNA library construction
kit. Complementary DNA was synthesized from mRNA using a
poly (dT) sequence with a NotI restriction site. SalI
linkers adapters were ligated to the blunt-ended cDNA
fragments followed by NotI digestion. The cDNA fragments
were directionally cloned into the NotI-SalI restriction
site of the pSPOR1 vector. The ligated cDNA fragments
were transformed into E.coli ElectroMax DH10B host cells.
This library was constructed in the laboratory of Dr. Lila
Vodkin by Anu Khanna at the University of Illinois at
Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

ORIGIN
Query Match 79.2%; Score 19.8; DB 12; Length 427;
Best Local Similarity 91.3%; Pred. No. 2.6e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAGTCAAAAACATATATGACTTA 24
||| ||||| ||||| ||||| |||||
Db 410 AACTCAAAAACATAGTACTTA 388

RESULT 7
BF596529/c
LOCUS
DEFINITION
BF596529.1 GI:11688853
433 bp mRNA linear EST 06-DEC-2001
su72e06.v1 Gm-cl055 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl055-227 5' similar to TR:Q95QU7 Q95QU7 F24P17.7 PROTEIN. ;,
mRNA sequence.
BF596529
Glycine max (soybean)
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 433)
Shoemaker,R., Keim,P., Vodkin,L., Expelding,J., Coryell,V.,
Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
Beck,C., Wylie,T., Underwood,K., Stentoe,M., Theising,B., Allen,M.,
Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,I., Cardenas,M.,
McCann,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Putative full length read
vector to vector length is 533 This clone is available through:
Reegen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL
35801 For further information call: (800)-533-4363 or contact via
email: cu@reegen.com
High quality sequence stop: 424.
Location/Qualifiers
1. 433
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl055-227"
/tissue_type="Mature seed pods, greenhouse grown"

/tissue_type="somatic embryos cultured on MSD 20"
/lab_host="DH10B"
/clone_lib="Gm-cl036"
/notes="Vector: pSPOR1; Site 1: NotI; Site 2: SalI; This
cDNA library was constructed from mRNA isolated from
somatic embryos (age ranging from 2 months to 9 months)
cultured on MSD 20. The library was prepared using the
Life Technologies psuperScript cDNA library construction
kit. Complementary DNA was synthesized from mRNA using a
poly (dT) sequence with a NotI restriction site. SalI
linkers adapters were ligated to the blunt-ended cDNA
fragments followed by NotI digestion. The cDNA fragments
were directionally cloned into the NotI-SalI restriction
site of the pSPOR1 vector. The ligated cDNA fragments
were transformed into E.coli ElectroMax DH10B host cells.
This library was constructed in the laboratory of Dr. Lila
Vodkin by Anu Khanna at the University of Illinois at
Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

ORIGIN
Query Match 79.2%; Score 19.8; DB 12; Length 427;
Best Local Similarity 91.3%; Pred. No. 2.6e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAGTCAAAAACATATATGACTTA 24
||| ||||| ||||| ||||| |||||
Db 410 AACTCAAAAACATAGTACTTA 388

RESULT 7
BF596529/c
LOCUS
DEFINITION
BF596529.1 GI:11688853
433 bp mRNA linear EST 06-DEC-2001
su72e06.v1 Gm-cl055 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl055-227 5' similar to TR:Q95QU7 Q95QU7 F24P17.7 PROTEIN. ;,
mRNA sequence.
BF596529
Glycine max (soybean)
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 433)
Shoemaker,R., Keim,P., Vodkin,L., Expelding,J., Coryell,V.,
Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
Beck,C., Wylie,T., Underwood,K., Stentoe,M., Theising,B., Allen,M.,
Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,I., Cardenas,M.,
McCann,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Putative full length read
vector to vector length is 533 This clone is available through:
Reegen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL
35801 For further information call: (800)-533-4363 or contact via
email: cu@reegen.com
High quality sequence stop: 424.
Location/Qualifiers
1. 433
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl055-227"
/tissue_type="Mature seed pods, greenhouse grown"
```

```
/lab_host="DH10B"
/clone_lib="Gm-cl055"
/notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from mRNA isolated
from mature seed pods of greenhouse grown plants prior to
senescence for the cultivar AP1. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with a XhoI restriction site. EcoRI
adapters were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA fragments were
directionally cloned into the EcoRI-XhoI restriction site
of the pBluescript vector. The ligated cDNA fragments
were transformed into DH10B host cells (GibcoBRL). This
library was constructed in the laboratory of Dr. Randy
Shoemaker."
```

ORIGIN

Query Match 79.2%; Score 19.8; DB 10; Length 433;
Best Local Similarity 91.3%; Pred. No. 2.6e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAGTCAAAAACATATATGACTTA 24
||| ||||| ||||| ||||| |||||
Db 391 AACTCAAAAACATAGTACTTA 369

RESULT 8
CC056262/c
LOCUS
DEFINITION
CC056262.1 GI:29475926
495 bp DNA linear GSS 02-APR-2003
SALK_101695.17.45.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_101695.17.45.x, genomic
survey sequence.
CC056262
CC056262.1 GI:29475926
GSS
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1 (bases 1 to 495)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
Shinn,P., Zimmermann,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA. This sequence lies within an annotated intron of Atg28260.
Class: TDNA tagged.
Location/Qualifiers
1. 495
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_101695.17.45.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/notes="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

FEATURES
source

ORIGIN

```

Query Match      79.2%; Score 19.8; DB 28; Length 495;
Best Local Similarity 91.3%; Pred. No. 2.5e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAAGTCAAAACATATATGACTT 23
    |||||
Db 224 TAAGTCAAAACATATATATCTT 202

RESULT 9
CD391412
LOCUS CD391412 666 bp mRNA linear EST 01-JUN-2003
DEFINITION Gm cki0293 Soybean induced by Salicylic Acid Glycine max cDNA 3',
            mRNA sequence.
ACCESSION CD391412
VERSION CD391412.1 GI:31306209
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max

REFERENCE 1 (bases 1 to 666)
AUTHORS Tian,A.-G., Wang,J., Cui,P., Han,Y.-J., Xu,H., Cong,L.-J.,
        Huang,X.-G., Wang,X.-L., Jiao,Y.-Z., Wang,B.-J., Wang,Y.-J.,
        Zhang,J.-S., Chen,S.-Y. and Yu,J.
        Soybean Expressed Sequence Tags Sequencing
        Unpublished (2003)
        Contact: Chen S.-Y
        Plant Biotechnology Laboratory
        Institute of Genetics and Developmental Biology, CAS, China
        Datun road, Beijing 100101, China
        Tel: 86-10-64886859
        Fax: 86-10-64873428
        Email: sychen@genetics.ac.cn
        Email: sychen@genetics.ac.cn
        Seq primer: T7 primer.
        Location/Qualifiers
            1..666
                /organism="Glycine max"
                /mol_type="mRNA"
                /cultivar="Kefeng 1"
                /db_xref="taxon:3847"
                /tissue_type="Seedlings"
                /dev_stage="two-week seedlings"
                /lab_host="XLI-Blue MRF, strain"
                /clone_lib="Soybean induced by Salicylic Acid"
                /note="Vector: pBluescript SK+; Site_1: EcoR I; Site 2:
                Xho I; The cDNA library was constructed by He, C-Y from
                mRNA isolated from two-week seedlings (cultivar Kefeng 1)
                treated by spraying 2.0mM salicylic acid for 24, 36, 48
                and 72 h. Complementary DNA was synthesized from mRNA
                using a primer consisting of a poly(dT) sequence with a
                XhoI restriction site. EcoRI adapters were ligated to the
                blunt-ended cDNA fragments followed by XhoI digestion. The
                cDNA fragments were directionally cloned into the
                EcoRI-XhoI restriction site of the pBluescript vector. The
                ligated cDNA fragments were transformed into XLI-Blue MRF,
                host cells (Stratagene)."
```

```

Query Match      79.2%; Score 19.8; DB 14; Length 666;
Best Local Similarity 91.3%; Pred. No. 2.4e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAGTCAAAACATATATGACTTA 24
    |||||
Db 293 AAGTCAAAACATATATGACTTA 315

RESULT 10
BM307747
```

```

LOCUS BM307747 233 bp mRNA linear EST 02-JAN-2002
DEFINITION Sak3d01.y1 Gm-cl075 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-cl075-4777 5', mRNA sequence.
ACCESSION BM307747
VERSION BM307747.1 GI:18039453
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
```

```

REFERENCE 1 (bases 1 to 233)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
        Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
        Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
        Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
        Schurr,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
        McCann,R., Waterston,R. and Wilson,K.
        Public Soybean EST project
        Unpublished (1999)
        Contact: Shoemaker R/Public Soybean EST Project
        Public Soybean EST Project
        Washington University School of Medicine
        4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
        Tel: 314 286 1800
        Fax: 314 286 1810
        Email: est@watson.wustl.edu
```

```

TITLE This clone is available through: ResGen, Invitrogen Corp. 2130
JOURNAL South Memorial Parkway Huntsville, AL 35801 For further information
COMMENT call: (800) 533-4363 or contact: cu@resgen.com web site:
        www.resgen.com
        Seq primer: -40RP from Gibco
        High quality sequence stop: 223.
        Location/Qualifiers
            1..233
                /organism="Glycine max"
                /mol_type="mRNA"
                /db_xref="taxon:3847"
                /clone="SOYBEAN CLONE ID: Gm-cl075-4777"
                /tissue_type="differentiating somatic embryos cultured on
                MSM6AC"
                /lab_host="DH10B"
                /clone_lib="Gm-cl075"
                /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
                XhoI; The cDNA library was constructed from mRNA isolated
                from differentiating somatic embryos cultured on MSM6AC.
                The library was prepared using the Stratagene pBluescript
                II SK(+) library construction kit. Complementary DNA was
                synthesized from mRNA using a primer consisting of a
                poly(dT) sequence with an XhoI restriction site. EcoRI
                adapters were ligated to the blunt-ended cDNA fragments
                followed by XhoI digestion. The cDNA fragments were
                directionally cloned into the EcoRI-XhoI restriction site
                of the pBluescript vector. The ligated cDNA fragments
                were transformed into E.coli ElectroMax DH10B host cells.
                Tissue culture and library construction were performed by
                Francoise Thibaud-Nissen and Anu Khana (Lila Vodkin lab,
                University of Illinois)."
```

```

FEATURES
source
    1..233
        /organism="Glycine max"
        /mol_type="mRNA"
        /db_xref="taxon:3847"
        /clone="SOYBEAN CLONE ID: Gm-cl075-4777"
        /tissue_type="differentiating somatic embryos cultured on
        MSM6AC"
        /lab_host="DH10B"
        /clone_lib="Gm-cl075"
        /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
        XhoI; The cDNA library was constructed from mRNA isolated
        from differentiating somatic embryos cultured on MSM6AC.
        The library was prepared using the Stratagene pBluescript
        II SK(+) library construction kit. Complementary DNA was
        synthesized from mRNA using a primer consisting of a
        poly(dT) sequence with an XhoI restriction site. EcoRI
        adapters were ligated to the blunt-ended cDNA fragments
        followed by XhoI digestion. The cDNA fragments were
        directionally cloned into the EcoRI-XhoI restriction site
        of the pBluescript vector. The ligated cDNA fragments
        were transformed into E.coli ElectroMax DH10B host cells.
        Tissue culture and library construction were performed by
        Francoise Thibaud-Nissen and Anu Khana (Lila Vodkin lab,
        University of Illinois)."
```

```

ORIGIN
Query Match      76.8%; Score 19.2; DB 12; Length 233;
Best Local Similarity 87.5%; Pred. No. 4.6e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAAGTCAAAACATATATGACTTA 24
    |||||
Db 36 TAAGTCAAAACATATATGACTTA 59

RESULT 11
AV538859
LOCUS AV538859 262 bp mRNA linear EST 07-SEP-2000
```

The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: <http://genome.sc.riken.go.jp/>
Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,
Masuura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and
Hayashizaki, Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 95 (7), 3455-3460 (1998)
Itoh, M., Kitsuai, T., Akiyama, Y., Shibata, K., Izawa, M., Kawai, S.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,
Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. *Genome Res.* 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp/>) for
further details.

```

Accession:
Location/Qualifiers
1.306
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="6330512K13"
/sex="male"
/tissue_type="medulla oblongata"
/dev_stage="adult"
/lab_host="PH10B"
/clone_lib="RIKEN full-length enriched, adult male medulla oblongata"

```

/note=*Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCAGTTATTTATTTATTTTCCTCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end: BamHI"

ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 306;
Best Local Similarity 87.5%; Pred. No. 4.4e+03;
Matches 21; Conservative 3; Indels 0; Gaps 0;

[illegible]

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Silurana.

REFERENCE AUTHORS TITLE JOURNAL COMMENT

1 (bases 1 to 371)
Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
Contact: Croning MDR

Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: TGAS080p20.gikt7
Sequencing primer: 5'

This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.
cDNA was oligo dT primed from sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli XL1-blue.

FEATURES source

Location/Qualifiers
1..371
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TGAS080p20"
/db_xref="taxon:8364"
/stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dT primed from sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 76.8%; Score 19.2; DB 13; Length 371;
Best Local Similarity 87.5%; Pred. No. 4.3e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAGTCAAAACATATATGACTTAA 25
DB 135 AATCCAAAATATATGACTTAA 158

RESULT 14

AQ986840/c
LOCUS
DEFINITION
RPCI-23-308D20-TV RPCI-23 Mus musculus genomic clone
RPCI-23-308D20, genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE

AQ986840 1 GI:6820045
Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 441)
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levine, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)

Other_GSSs: RPCI-23-308D20.TJ
Contact: Shaying Zhao

Department of Eukaryotic Genomics
The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 308 row: D column: 20
Seq primer: T7
Class: BAC ends.

FEATURES source

Location/Qualifiers
1..441
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clones="RPCI-23-308D20"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI; female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN

Query Match 76.8%; Score 19.2; DB 28; Length 441;
Best Local Similarity 87.5%; Pred. No. 4.1e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAGTCAAAACATATATGACTTAA 25
DB 321 AAGTGAAGAAACATATTTGACTTAA 298

RESULT 15

AW458707/c
LOCUS
DEFINITION
SH2f03.y1 Gm-cl016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl016-4566 5', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE

AW458707 1 GI:7028924
EST.
Glycine max (soybean)

ORGANISM

Glycine max (soybean)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

1 (bases 1 to 481)

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvelli, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, I., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, J., Person, B., Swaller, J., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

Public Soybean EST Project
Unpublished (1999)

Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information

call: (800)-533-4363 or contact via email: ccu@resgen.com
Insert Length: 1046 Std Error: 0.00

Seq primer: -40RP from Gibco
High quality sequence stop: 411.

FEATURES Location/Qualifiers

```
source
1. 481
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl016-4566"
/tissue_type="immature flowers of field grown plants"
/lab_host="Xli0-Gold"
/clone_lib="Gm-cl016"
/note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2:
XhoI; This cDNA library was constructed from mRNA isolated
from immature flowers of field grown plants. The cDNA isolated
library was prepared using the Stratagene pBluescript II
XR library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI adapters
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into Xli0-Gold host cells. This library was
constructed by Dr. Randy Shoemaker and Dr. John
Erpelding."
```

ORIGIN

```
Query Match      76.8%; Score 19.2; DB 10; Length 481;
Best Local Similarity 87.5%; Pred. No. 4.1e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  TAAATCAAAACATATATGACTTAA 24
      |||||
Db      50  TAAATCGAAATATATGACTTAA 27
```

Search completed: May 26, 2004, 22:27:15
Job time : 1732.16 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 15:22:28 ; Search time 164.244 Seconds
(without alignments)
594,900 Million cell updates/sec

Title: US-10-676-299-7

Perfect score: 23

Sequence: 1 ttaatcatatgcgttttgggtta 23

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	100.0	23	8	ACD28588 E. coli A
2	23	100.0	30	8	ACD28584 E. coli A
3	21	91.3	21	8	ACD28589 E. coli A
4	21	91.3	28	8	ACD28585 E. coli A
5	18.8	81.7	3727	7	ADA53054 Human cod
6	18.8	81.7	26923	4	AAP28554 Genomic f
7	18.2	79.1	343	6	ABL7877 Human ova
8	18.2	79.1	1026	6	ABN92532 Staphyloc
9	17.8	77.4	2322	7	ABX11841 Bacillus
10	17.8	77.4	2475	9	ADE71408 Bacillus
11	17.8	77.4	2783	6	AAI69288 Bacillus
12	17.8	77.4	2923	1	AAAT70651 Sequence
13	17.8	77.4	3150	7	ABZ57796 Bacillus
14	17.8	77.4	3199	3	AA97493 Bacillus
15	17.8	77.4	3332	2	ABZ57797 Bacillus
16	17.8	77.4	3463	2	AAQ27180 Bacillus
17	17.8	77.4	10717	6	ABL33694 Human imm
18	17.8	77.4	10717	6	ABN80210 Human che
19	17.8	77.4	110000	2	AAAT42063_12
20	17.4	75.7	1371	7	ACA44727 Prokaryot
21	17.4	74.8	627	5	AA05443 Mammalian
22	17.2	74.8	639	5	ABQ42337 Oligonuel
23	17.2	74.8	639	6	ABQ42336 Oligonuel

ALIGNMENTS

RESULT 1

ACD28588
ID ACD28588 standard; DNA; 23 BP.

XX
AC ACD28588;

XX
DT 10-OCT-2003 (first entry)

DE
DB E. coli ArsrR binding oligonucleotide PLASSIT.

XX
XX ArsrR; arsenic resistance operon; biosensor; ss; arsenic.

XX
XX Escherichia coli.

XX
PN US2003096275-A1.

XX
PD 22-MAY-2003.

XX
PF 15-AUG-2002; 2002US-00222952.

XX
PR 20-AUG-2001; 2001US-0313714P.

XX
PA (LATIN/) LAING L G.

XX
PI Laing LG;

XX
DR WPI; 2003-576876/54.

XX
PT New system comprising isolated protein and nucleic acid, and a detection system that indicates a change in binding of the protein to the nucleic acid in the presence of the analyte, useful for detecting the presence of an analyte in a sample.

XX
PS Claim 35; Page 15; 36pp; English.

XX
CC The invention relates to a new system (biosensor) for detecting the presence of analyte in a sample comprising: (1) an isolated protein that specifically binds the analyte; (2) an isolated nucleic acid containing a specific binding sequence that is bound specifically by the protein; and (3) a detection system that indicates a change in binding of the protein to the nucleic acid in the presence of the analyte. Also included are a biosensor device for detecting the presence of an analyte in a sample, detecting the presence of an analyte in a sample, and an ArsrR (encoded by part of the arsenic resistance operon of E. coli) protein comprising an amino acid sequence that is at least 90% identical to amino acids 1-97 of the ArsrR sequence appearing as AB063440 binding to a nucleic acid

CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the top strand of an oligonucleotide which binds to a
 CC plasmid expressed ArsR protein and is used in the biosensor of the
 CC invention

SQ Sequence 23 BP; 5 A; 2 C; 4 G; 12 T; 0 U; 0 Other;

Query Match 100.0%; Score 23; DB 8; Length 23;

Best Local Similarity 100.0%; Pred. No. 1.6;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTAATCATATGCGTTTGGTTA 23

Db 1 TTAATCATATGCGTTTGGTTA 23

RESULT 2

ACD28584

ID ACD28584 standard; DNA; 30 BP.

AC ACD28584;

DT 10-OCT-2003 (first entry)

E. coli ArsR binding oligonucleotide PLASL17.

ArsR; arsenic resistance operon; biosensor; ss; arsenic.

Escherichia coli.

US2003096275-A1.

22-MAY-2003.

15-AUG-2002; 2002US-00222952.

20-AUG-2001; 2001US-0313714P.

(LAIN/) LAING L G.

Laing LG;

WPI; 2003-576876/54.

New system comprising isolated protein and nucleic acid, and a detection
 PT system that indicates a change in binding of the protein to the nucleic
 PT acid in the presence of the analyte, useful for detecting the presence of
 PT analyte in a sample.

Claim 35; Page 15; 36pp; English.

The invention relates to a new system (biosensor) for detecting the
 CC presence of analyte in a sample comprising: (1) an isolated protein that
 CC specifically binds the analyte; (2) an isolated nucleic acid containing a
 CC specific binding sequence that is bound specifically by the protein; and
 CC (3) a detection system that indicates a change in binding of the protein
 CC to the nucleic acid in the presence of the analyte. Also included are a
 CC biosensor device for detecting the presence of an analyte in a sample,
 CC detecting the presence of an analyte in a sample, and an ArsR (encoded by
 CC part of the arsenic resistance operon of E. coli) protein comprising an
 CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
 CC the ArsR sequence appearing as AB063440 binding to a nucleic acid
 CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the top strand of an oligonucleotide which binds to a
 CC plasmid expressed ArsR protein and is used in the biosensor of the
 CC invention

SQ Sequence 30 BP; 5 A; 2 C; 7 G; 16 T; 0 U; 0 Other;

Query Match

Best Local Similarity 100.0%; Score 23; DB 8; Length 30;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTAATCATATGCGTTTGGTTA 23

Db 1 TTAATCATATGCGTTTGGTTA 23

RESULT 3

ACD28589/c

ID ACD28589 standard; DNA; 21 BP.

AC ACD28589;

DT 10-OCT-2003 (first entry)

E. coli ArsR binding oligonucleotide PLASSIB.

ArsR; arsenic resistance operon; biosensor; ss; arsenic.

Escherichia coli.

US2003096275-A1.

22-MAY-2003.

15-AUG-2002; 2002US-00222952.

20-AUG-2001; 2001US-0313714P.

(LAIN/) LAING L G.

Laing LG;

WPI; 2003-576876/54.

New system comprising isolated protein and nucleic acid, and a detection
 PT system that indicates a change in binding of the protein to the nucleic
 PT acid in the presence of the analyte, useful for detecting the presence of
 PT analyte in a sample.

Claim 35; Page 15; 36pp; English.

The invention relates to a new system (biosensor) for detecting the
 CC presence of analyte in a sample comprising: (1) an isolated protein that
 CC specifically binds the analyte; (2) an isolated nucleic acid containing a
 CC specific binding sequence that is bound specifically by the protein; and
 CC (3) a detection system that indicates a change in binding of the protein
 CC to the nucleic acid in the presence of the analyte. Also included are a
 CC biosensor device for detecting the presence of an analyte in a sample,
 CC detecting the presence of an analyte in a sample, and an ArsR (encoded by
 CC part of the arsenic resistance operon of E. coli) protein comprising an
 CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
 CC the ArsR sequence appearing as AB063440 binding to a nucleic acid
 CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the bottom strand of an oligonucleotide which binds
 CC to a plasmid expressed ArsR protein and is used in the biosensor of the
 CC invention

SQ Sequence 21 BP; 10 A; 4 C; 2 G; 5 T; 0 U; 0 Other;

Query Match

Best Local Similarity 91.3%; Score 21; DB 8; Length 21;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 AATCATATGCGTTTGGTTA 23

Db 21 AATCATATGCGTTTGGTTA 1

RESULT 4

ACD28585/c

ID ACD28585 standard; DNA; 28 BP.

XX ACD28585;
 XX 10-OCT-2003 (first entry)
 XX E. coli ArsR binding oligonucleotide PLASL1B.
 XX ArsR; arsenic resistance operon; biosensor; ss; arsenic.
 XX Escherichia coli.
 XX US2003096275-A1.
 XX 22-MAY-2003.
 XX 15-AUG-2002; 2002US-00222952.
 XX 20-AUG-2001; 2001US-0313714P.
 XX (LAIN/) LAING L G.
 XX Laing LG;
 XX WPI; 2003-576876/54.
 XX New system comprising isolated protein and nucleic acid, and a detection
 PT system that indicates a change in binding of the protein to the nucleic
 PT acid in the presence of the analyte, useful for detecting the presence of
 PT analyte in a sample.
 XX Claim 35; Page 15; 36pp; English.
 XX The invention relates to a new system (biosensor) for detecting the
 CC presence of analyte in a sample comprising: (1) an isolated protein that
 CC specifically binds the analyte; (2) an isolated nucleic acid containing a
 CC specific binding sequence that is bound specifically by the protein; and
 CC (3) a detection system that indicates a change in binding of the protein
 CC to the nucleic acid in the presence of the analyte. Also included are a
 CC biosensor device for detecting the presence of an analyte in a sample,
 CC detecting the presence of an analyte in a sample, and an ArsR (encoded by
 CC part of the arsenic resistance operon of E. coli) protein comprising an
 CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
 CC the ArsR sequence appearing as AB063440 binding to a nucleic acid
 CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the bottom strand of an oligonucleotide which binds
 CC to a plasmid expressed ArsR protein and is used in the biosensor of the
 CC invention.
 XX Sequence 28 BP; 14 A; 7 C; 2 G; 5 T; 0 U; 0 Other;
 XX
 Query Match 91.3%; Score 21; DB 8; Length 28;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 AATCATATCGGTTTGGTTA 23
 DB 28 AATCATATCGGTTTGGTTA 8
 RESULT 5
 ID ADA53054/c
 ID ADA53054 standard; cDNA; 3727 BP.
 XX ADA53054;
 XX 20-NOV-2003 (first entry)
 XX Human coding sequence, SEQ ID 622.
 XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
 KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
 KW inflammatory disease; osteoporosis; neurological disease; gene; ss.

XX Homo sapiens.
 XX EP1293569-A2.
 XX 19-MAR-2003.
 XX 21-MAR-2002; 2002EP-00006586.
 XX 14-SEP-2001; 2001JP-00328381.
 XX 24-JAN-2002; 2002US-0350435P.
 XX (HELI-) HELIX RES INST.
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
 XX WPI; 2003-395539/38.
 XX P-FSDB; ADA54693.
 XX New polynucleotides encoding full-length polypeptides, e.g. secretory
 PT and/or membrane proteins, useful for developing medicines for diseases in
 PT which the gene is involved, or as target molecules for gene therapy.
 XX Claim 1; SEQ ID NO 622; 205pp; English.
 XX The present invention relates to novel human secretory or membrane
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
 CC ADA54071). The coding sequences are useful in the gene therapy of
 CC diseases caused by abnormalities of the proteins, e.g. cancer,
 CC inflammatory diseases, osteoporosis or neurological disease.
 XX Sequence 3727 BP; 927 A; 839 C; 1016 G; 945 T; 0 U; 0 Other;
 XX
 Query Match 81.7%; Score 18.8; DB 7; Length 3727;
 Best Local Similarity 90.9%; Pred. No. 1.1e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 TTAATCATATCGGTTTGGTT 22
 DB 2104 TTTATCATATCGGTTTGGTT 2083
 RESULT 6
 ID AAF28554/c
 ID AAF28554 standard; DNA; 269223 BP.
 XX AAF28554;
 XX 04-APR-2001 (first entry)
 XX Genomic fragment #41.
 XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 KW bronchopulmonary; endocarditis; meningitis; ss.
 XX Moraxella catarrhalis.
 XX WO200078968-A2.
 XX 28-DEC-2000.
 XX 16-JUN-2000; 2000WO-US016649.
 XX 18-JUN-1999; 99US-0140121P.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Lagace RE, Patterson C, Berg KL;
 XX WPI; 2001-041427/05.

XX Genomic library for identifying diagnostic and therapeutic compositions,
PT and for identifying virulence factors, regulatory elements and drug
PT targets, comprises Moraxella catarrhalis nucleic acids.
XX
XX
PS Claim 1; Page 486-545; 545pp; English.
XX
XX The present invention relates to a Moraxella catarrhalis genomic library
CC comprising of a combination of 41 nucleic acid molecules (see AAF28514-
CC AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis
XX
SQ Sequence 269223 BP; 77067 A; 56596 C; 57380 G; 78180 T; 0 U; 0 Other;
Query Match 81.7%; Score 18.8; DB 4; Length 269223;
Best Local Similarity 90.9%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 TTAATCATATGCGTTTGGTT 22
Db 142085 TTAATGATATGCGTTTGGCT 142064
RESULT 7
ABL87877
ID ABL87877 standard; cDNA; 343 BP.
XX
AC ABL87877;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human ovarian cancer related cDNA clone SEQ ID NO:10855.
XX
KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200192581-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US017756.
XX
PR 26-MAY-2000; 2000US-0207484P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Algate PA, Harlocker SL, Jones R;
XX
DR WPI; 2002-122075/16.
XX
PT Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide.
XX
PS Claim 1; SEQ ID NO 10855; 489pp; English.
XX
XX The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC from the 19912 nucleotide sequences as given in ABL77023 to ABL87934,
CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC or antigen presenting cells that express (II). (I) has cytostatic
CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC detecting ovarian cancer in a patient's biological sample preferably
CC serum or ovarian tissue. The method comprises contacting a biological

CC sample from a patient with (IV), detecting the amount of polynucleotide
CC hybridising to (IV) and comparing the amount to a predetermined cutoff
CC value and thereby detecting ovarian cancer in the patient, where the
CC amount of polynucleotide hybridising to (IV) is detected preferably by
CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC useful for stimulating and/or expanding T cells specific for an ovarian
CC tumour protein comprising contacting T cells with (III) or (II). (III) is
CC useful in design and preparation of ribozyme molecules for inhibiting
CC expression of the tumour polypeptides and proteins in tumour cells; and
CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC library using well known techniques
XX
SQ Sequence 343 BP; 117 A; 52 C; 50 G; 124 T; 0 U; 0 Other;
Query Match 79.1%; Score 18.2; DB 6; Length 343;
Best Local Similarity 87.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 TTAATCATATGCGTTTGGTTA 23
Db 78 TTAATCAGATGCCCTTTAGTTA 100
RESULT 8
ABN92532/c
ID ABN92532 standard; DNA; 1026 BP.
XX
AC ABN92532;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1995.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX antibacterial; gene therapy; gene; ds.
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-00134001.
XX
PR 14-AUG-1997; 97US-0055779P.
XX
PR 08-NOV-1997; 97US-0064964P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2002-381255/41.
XX
DR P-PSDB; ABP39987.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermis
PT polypeptide, useful for diagnosing and treating bacterial infections.
XX
PS Disclosure; SEQ ID NO 1995; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences can
CC also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life cycle
CC or inhibit S. epidermidis infection. N.B. The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the USPTO web site
XX
SQ Sequence 1026 BP; 369 A; 124 C; 231 G; 302 T; 0 U; 0 Other;
Query Match 79.1%; Score 18.2; DB 6; Length 1026;
XX

Best Local Similarity 87.0%; Pred. No. 2a+02; Mismatches 3; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY 1 TTAATCATATGCTTTTGGTGA 23
|||||
Db 624 TTAATCATATGCTTTTGGTGA 602

RESULT 9
ABX11841/c
ID ABX11841 standard; DNA; 2322 BP.
XX
AC ABX11841;
XX
DT 27-OCT-2003 (revised)
DT 07-MAY-2003 (first entry)
XX
DE Bacillus sp. DNA encoding endo-beta-1,4-glucanase.
XX
KW Gene; ds; endo-beta-1,4-glucanase; detergent; textile finishing process;
KW oil industry; biomass degradation; laundry; stone washing;
KW pulp processing; animal feed.
XX
OS Bacillus sp; AA349 strain DSM 12648.
XX
FH Key Location/Qualifiers
FT CDS 1..2322
FT /*tag= a
FT /EC_number= "EC 3.2.1.4"
FT /product= "Endo-beta-1,4-glucanase"
FT /partial
FT /note= "No start codon shown"
XX
PN WO200299091-A2.
XX
PD 12-DEC-2002.
XX
PF 06-JUN-2002; 2002WO-DK000381.
XX
PR 06-JUN-2001; 2001DK-00000879.
XX
PA (NOVO) NOVOZYMES AS.
XX
PI Outtrup H, Schuelein M, Eskelund MB, Gibson K;
XX
DR WPI; 2003-256232/25.
DR P-PSDB; ABG76403.
XX
PT New enzyme exhibiting endo-beta-1,4-glucanase activity, useful in
PT detergent compositions, oil industry, textile finishing processes, biomass
PT degradation, laundry, and stone washing.
XX
PS Claim 5; Page 42-45; 51pp; English.
XX
CC The invention relates to an enzyme exhibiting endo-beta-1,4-glucanase
CC activity (EC 3.2.1.4), comprising: (a) a polypeptide encoded by the DNA
CC sequence appearing as ABX11841; (b) a polypeptide produced by culturing a
CC cell comprising the DNA sequence under conditions where the DNA sequence
CC is expressed; (c) an endo-beta-1,4-glucanase enzyme having at least 97%
CC sequence identity to the amino acid sequence appearing as ABG76403; or
CC (d) a polypeptide having endo-beta-1,4-glucanase activity that is encoded
CC by a polynucleotide that hybridizes to the DNA under hybridisation
CC conditions comprising 5X SSC at 45 plusOC and washing conditions
CC comprising 2X SSC at 60 plusOC. Also included are the encoding DNA
CC sequence, a polynucleotide construct comprising any of the DNA sequence,
CC an expression vector (comprising the following operably linked elements:
CC a transcription promoter, a DNA segment encoding the enzyme and a
CC transcription terminator), a cultured cell comprising the vector and
CC expressing the enzyme, a method for degradation of cellulose-containing
CC biomass that is treated with the enzyme or enzyme composition cited above
CC and a hybrid endo-glucanase (exhibiting endo- beta-1,4-glucanase activity
CC comprising the cellulase binding domain, CBD, of the enzyme , and a
CC catalytic domain (CAD) from sources other than Bacillus sp. AA349 strain

DSM12648). The enzymes are useful in detergent composition, textile finishing processes, oil industry, biomass degradation, laundry and stone washing. The invention provides enzymes having substantial beta-1,4-glucanase activity under slightly acid to alkaline conditions and improved performance in pulp processing, textile treatment, laundry processes, extraction processes or in animal feed. The present sequence encodes the endo-beta-1,4-glucanase. (Updated on 27-OCT-2003 to standardise OS field)

XX Sequence 2322 BP; 774 A; 374 C; 546 G; 627 T; 0 U; 1 Other;
SQ

Query Match 77.4%; Score 17.8; DS 7; Length 2322;
Best Local Similarity 90.5%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTAATCATATGCTTTTGGT 21
|||||
Db 2015 TTAATCGTATACGTTTGGT 1995

RESULT 10
ADE71408/c
ID ADE71408 standard; DNA; 2475 BP.
XX
AC ADE71408;
XX
DT 29-JAN-2004 (first entry)
XX
DE Bacillus sp. KSM-S237 alkaline cellulase Egl-237-encoding Gene.
XX
KW Alkaline cellulase; Egl-237; wild-type; mutant enzyme; detergent;
KW laundry; gene; ds.
XX
OS Bacillus sp. KSM-S237.
XX
FH Key Location/Qualifiers
FT CDS 1..2475
FT /*tag= a
FT /product= "Alkaline cellulase Egl-237"
XX
PN WO2003091422-A1.
XX
PD 06-NOV-2003.
XX
PF 25-APR-2003; 2003WO-JP005371.
XX
PR 25-APR-2002; 2002JP-00124474.
XX
PA (KAOS) KAO CORP.
XX
PI Hakamada Y, Ozawa T, Kobayashi T;
XX
DR WPI; 2003-854397/79.
DR P-PSDB; ADE71407.
XX
PS Mutated alkaline cellulase for use as an enzyme for detergents is
XX produced by deleting one or more amino acid residue groups from the 343-
XX to 373-positions in SEQ ID NO:1 and then inserting a peptide into the
XX deletion site.
XX
XX Disclosure; SEQ ID NO 2; 31pp; Japanese.
XX
CC The invention relates to a mutant alkaline cellulase derived from the
CC Bacillus sp. KSM-S237 alkaline cellulase Egl-237 (ADE71407). The mutant
CC enzyme is created by deleting one or more amino acid residues between
CC residues 343-373 of the wild-type enzyme, and then inserting a 2-15
CC residue peptide into the deletion site. The invention also encompasses a
CC gene encoding a mutant alkaline cellulase of the invention, and vectors
CC and host cells comprising a mutant alkaline cellulase-encoding gene. The
CC mutant alkaline cellulases of the invention have an optimum pH which is
CC very close to the pH of laundry water (around pH 10.5) and are therefore
CC useful as enzymes for detergents. The present sequence represents the
CC gene encoding wild-type Bacillus sp. KSM-S237 alkaline cellulase Egl-237.

XX SQ Sequence 2475 BP; 838 A; 394 C; 574 G; 669 T; 0 U; 0 Other;
 Query Match 77.4%; Score 17.8; DB 9; Length 2475;
 Best Local Similarity 90.5%; Pred. No. 2.9e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTAATCATATCGTTTGGT 21
 ||||| ||||| ||||| ||||| |||||
 Db 2102 TTAATCGTATACGTTTGGT 2082

RESULT 11
 AA169288/c
 ID AA169288 standard; DNA; 2783 BP.
 AC
 XX AA169288;
 XX
 XX 08-FEB-2002 (first entry)
 XX
 XX Bacillus sp alkaline cellulase N131b encoding DNA.
 DE
 XX Alkaline cellulase; N131b; textile; detergent; treating agent; ds.
 KW
 XX Bacillus sp.
 XX
 XX Key Location/Qualifiers
 FH 293..2734
 FT CDS /*tag= a
 FT /*tag= b
 FT /*tag= c
 FT /*tag= d
 FT /*tag= e
 FT JP2001231569-A.
 PN
 XX
 XX 28-AUG-2001.
 XX
 XX 24-FEB-2000; 2000JP-00047237.
 XX
 XX 24-FEB-2000; 2000JP-00047237.
 PR
 XX (KAOS) KAO CORP.
 PA
 XX WPI; 2002-029359/04.
 XX
 XX P-PSDB; AAG50267.
 DR
 XX
 XX Alkaline cellulase gene useful for the preparation of an alkaline
 PT cellulase useful as a textile detergent and a textile treating agent.
 PT
 XX Example 6; Page 15-18; 22pp; Japanese.
 PS
 XX This invention describes a novel alkaline cellulase gene from a Bacillus
 CC sp. The alkaline cellulase gene is used for the preparation of an
 CC alkaline cellulase useful as a textile detergent and a textile treating
 CC agent. This sequence encodes the Bacillus sp. alkaline cellulase N131b
 CC described in the method of the invention
 XX
 XX SQ Sequence 2783 BP; 925 A; 468 C; 635 G; 755 T; 0 U; 0 Other;
 Query Match 77.4%; Score 17.8; DB 6; Length 2783;
 Best Local Similarity 90.5%; Pred. No. 3e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTAATCATATCGTTTGGT 21
 ||||| ||||| ||||| ||||| |||||
 Db 2361 TTAATCGTATACGTTTGGT 2341

RESULT 12
 AA70651/c
 ID AA70651 standard; DNA; 2923 BP.
 AC
 XX AA70651;
 XX

DT 24-OCT-2003 (revised)
 DT 20-JAN-1991 (first entry)
 XX
 XX Sequence of cellulase gene derived from Bacillus sp. No. 1139.
 DE
 XX Enzyme; cellotriose; cellotetrose; hydrolysis; ss.
 KW
 XX Bacillus sp; No. 1139.
 OS
 XX Key Location/Qualifiers
 FH 193..199
 FT RBS /*tag= a
 FT CDS 203..292
 FT /*tag= b
 FT mat_peptide 293..2605
 FT /*tag= c
 FT misc_structure 2690..2706
 FT /*tag= d
 FT misc_structure 2709..2726
 FT /*tag= e
 FT JP62232386-A.
 PN
 XX
 XX 12-OCT-1987.
 XX
 XX 02-APR-1986; 86JP-00076285.
 XX
 XX 02-APR-1986; 86JP-00076285.
 PR
 XX (RIKA) RIKAGAKU KENKYUSHO.
 PA
 XX WPI; 1987-325245/46.
 XX
 XX P-PSDB; AAP70420.
 DR
 XX DNA sequence coding cellulase gene - is derived from Bacillus sp. no 1139
 PT z and can hydrolyse cello-triose and cello-tetrose.
 PT
 XX Disclosure; Fig 3-1 - 3-3; 11pp; Japanese.
 PS
 XX By using this DNA sequence cellulase can be produced. This cellulase
 CC hydrolyses cellotriose or cellotetrose and does not hydrolyse cellobiose.
 CC Molecular weight is approx. 92K dalton and optimum pH is 9. (Updated on
 CC 24-OCT-2003 to standardise OS field)
 XX
 XX SQ Sequence 2923 BP; 992 A; 455 C; 654 G; 822 T; 0 U; 0 Other;
 Query Match 77.4%; Score 17.8; DB 1; Length 2923;
 Best Local Similarity 90.5%; Pred. No. 3e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTAATCATATCGTTTGGT 21
 ||||| ||||| ||||| ||||| |||||
 Db 2298 TTAATCGTATACGTTTGGT 2278

RESULT 13
 AB257796/c
 ID AB257796 standard; DNA; 3150 BP.
 XX
 XX AC AB257796;
 XX
 XX 16-APR-2003 (first entry)
 DT
 XX Bacillus sp. KSM-S237 sporulation-associated gene #1, SEQ ID NO:1.
 DE
 XX Sporulation associated gene; deactivation; deletion;
 KW sporulation suppression; sigE; sigF; spoIIE; spoIISB; sigG; spoIVCB;
 KW spoIIC; recombinant protein production; gene; ds.
 XX
 XX Bacillus sp. KSM-S237.
 OS
 XX Key Location/Qualifiers
 FH 573..3047
 FT CDS

```
FT FT /*tag= a
FT FT /product= "Sporulation-associated protein #1"
FT FT sig_peptide 573..659
FT FT /*tag= b
FT FT mat_peptide 660..3044
FT FT /*tag= c
FT FT /product= "Mature sporulation-associated protein #1"
XX XX
XX XX W0200297064-A1.
XX XX
XX XX 05-DEC-2002.
XX XX
XX XX 28-MAY-2002; 2002WO-JP005151.
XX XX
XX XX 29-MAY-2001; 2001JP-00160520.
XX XX
XX XX (KAOS ) KAO CORP.
XX XX
XX XX Sawada K, Endo K, Ozawa T, Tohata M, Ozaki K;
XX XX WPI; 2003-140463/13.
XX XX P-PSDB; ABP58992.
XX XX
XX XX Bacillus strain having genes associated with sporulation inactivated for
XX XX use as hosts in more efficient production of recombinant proteins.
XX XX
XX XX Disclosure; Page 14-19; 29pp; Japanese.
XX XX
XX XX The invention relates to microorganisms (particularly Bacillus species)
XX XX in which one or more sporulation-associated genes active in the medium to
XX XX latter stages of sporulation have been deactivated or deleted. Such genes
XX XX include the sigE, sigF, spoIIE, spoIIS, sigG, spoIVCB and spoIIC genes.
XX XX The invention also relates to the recombinant production of a protein
XX XX using the microorganism of the invention. As the microorganisms of the
XX XX invention are incapable of producing spores, recombinant protein
XX XX production is more efficient, with reduced formation of by-products,
XX XX reduced energy loss, reduced consumption of culture medium, and increased
XX XX speed of target protein production. In addition, protein production can
XX XX be efficiently carried out over a longer period of time than is currently
XX XX the case. The microorganisms can be used to recombinantly produce target
XX XX proteins for use in a variety of applications, such as in foodstuffs, a
XX XX drugs, cleaning agents and toiletries. The present sequence represents a
XX XX Bacillus sp. KSM-S237 sporulation-associated gene related to the
XX XX invention. Note: The present sequence is given in the sequence listing,
XX XX but is not further referred to in the specification
XX XX
XX XX Sequence 3150 BP; 1073 A; 474 C; 695 G; 908 T; 0 U; 0 Other;
XX XX
XX XX Query Match 77.4%; Score 17.8; DB 7; Length 3150;
XX XX Best Local Similarity 90.5%; Pred. No. 3e+02;
XX XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX XX
XX XX QY 1 TTAATCATATGCGTTTTCGT 21
XX XX ||||| ||||| ||||| |||||
XX XX DB 2674 TTAATCGTATACGTTTTCGT 2654
XX XX
XX XX RESULT 14
XX XX AAA97493/c
XX XX ID AAA97493 standard; DNA; 3189 BP.
XX XX
XX XX AC AAA97493;
XX XX
XX XX 29-JAN-2001 (first entry)
XX XX
XX XX Bacillus sp. KSM-S237 heat resistant alkaline cellulase DNA.
XX XX
XX XX Heat resistant alkaline cellulase; detergent; cleaning application; ds.
XX XX
XX XX Bacillus sp. KSM-S237.
XX XX
XX XX JP2000210081-A.
XX XX
```

```
PD 02-AUG-2000.
XX XX
XX XX 21-JAN-1999; 99JP-00013049.
XX XX
XX XX 21-JAN-1999; 99JP-00013049.
XX XX
XX XX (KAOS ) KAO CORP.
XX XX
XX XX WPI; 2000-596504/57.
XX XX P-PSDB; AAB23180.
XX XX
XX XX A heat resistant alkaline cellulase gene.
XX XX
XX XX Claim 3; Page 6-10; 10pp; Japanese.
XX XX
XX XX The invention relates to a novel heat resistant alkaline cellulase
XX XX (AAB23180) from Bacillus sp. KSM-S237, and to the gene encoding it
XX XX (AAA97493). The invention also encompasses variants of the heat resistant
XX XX alkaline cellulase in which at least one amino acid residue is deleted,
XX XX replaced or inserted. The Bacillus sp. KSM-S237 heat resistant alkaline
XX XX cellulase gene was isolated using the Bacillus sp. No. 1139 alkaline
XX XX cellulase gene-derived PCR primers AAA97494-A97495. The Bacillus sp. KSM-
XX XX S237 heat resistant alkaline cellulase gene can be used for the
XX XX recombinant production of heat resistant alkaline cellulase which can be
XX XX formulated in detergent for cleaning applications. The present sequence
XX XX represents the Bacillus sp. KSM-S237 heat resistant alkaline cellulase
XX XX gene
XX XX
XX XX Sequence 3189 BP; 1088 A; 476 C; 704 G; 921 T; 0 U; 0 Other;
XX XX
XX XX Query Match 77.4%; Score 17.8; DB 3; Length 3189;
XX XX Best Local Similarity 90.5%; Pred. No. 3e+02;
XX XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX XX
XX XX QY 1 TTAATCATATGCGTTTTCGT 21
XX XX ||||| ||||| ||||| |||||
XX XX DB 2674 TTAATCGTATACGTTTTCGT 2654
XX XX
XX XX RESULT 15
XX XX ABZ57797/c
XX XX ID ABZ57797 standard; DNA; 3332 BP.
XX XX
XX XX AC ABZ57797;
XX XX
XX XX 16-APR-2003 (first entry)
XX XX
XX XX Bacillus sp. KSM-64 sporulation-associated gene #2, SEQ ID NO:2.
XX XX
XX XX Sporulation associated gene; deactivation; deletion;
XX XX sporulation suppression; sigE; sigF; spoIIE; spoIISB; sigG; spoIVCB;
XX XX spoIIC; recombinant protein production; gene; ds.
XX XX
XX XX Bacillus sp. KSM-64.
XX XX
XX XX Key Location/Qualifiers
XX XX CDS 610..3078
XX XX /*tag= a
XX XX /product= "Sporulation-associated protein #2"
XX XX
XX XX sig_peptide 610..696
XX XX /*tag= b
XX XX mat_peptide 697..3075
XX XX /*tag= c
XX XX /product= "Mature sporulation-associated protein #2"
XX XX
XX XX W0200297064-A1.
XX XX
XX XX 05-DEC-2002.
XX XX
XX XX 28-MAY-2002; 2002WO-JP005151.
XX XX
XX XX 29-MAY-2001; 2001JP-00160520.
XX XX
```

Search completed: May 26, 2004, 17:50:10
Job time : 167.244 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 15:22:28 ; Search time 178.526 Seconds
(without alignments)
594,900 Million cell updates/sec

Title: US-10-676-299-10

Perfect score: 25

Sequence: 1 taagtcaaaaacatatgacttaa 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001as:*

5: Geneseq2001bs:*

6: Geneseq2002as:*

7: Geneseq2003as:*

8: Geneseq2003bs:*

9: Geneseq2003cs:*

10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	8	ACD28591 E. coli A
C 2	25	100.0	25	8	ACD28590 E. coli A
3	25	100.0	40	8	ACD28587 E. coli A
C 4	25	100.0	42	8	ACD28586 E. coli A
5	25	100.0	401	7	ACal15469 Prokaryot
C 6	25	100.0	2182	5	Aas82555 DNA encod
7	20.2	80.8	2000	6	Abx15842 Arabidops
C 8	20.2	80.8	5641	6	Ab133397 Human imm
9	19.8	79.2	62782	8	Aad58282 Human tum
10	19.8	79.2	62782	8	Aad58281 Human tum
11	19.8	79.2	94191	9	Continuation (4 of
12	19.8	79.2	226475	8	Aad58279 Human tum
C 13	19.2	76.8	1293	3	Aac36413 Arabidops
14	18.3	75.2	110000	6	Continuation (3 of
15	18.6	74.4	561	5	Abal3700 Human ner
16	18.6	74.4	668	5	Abal19908 Human ner
17	18.6	74.4	668	5	Abal20194 Human ner
18	18.6	74.4	379	5	Abal20198 Human ner
19	18.6	74.4	1772	5	Abal19910 Human ner
20	18.6	74.4	1772	5	Abal20197 Human ner
21	18.6	74.4	1772	5	Abal18947 Human ner
22	18.6	74.4	1772	5	Abal20195 Human ner
23	18.6	74.4	1772	5	Abal19911 Human ner

24 18.6 74.4 1772 5 ABA18949 Human ner

25 18.6 74.4 2585 5 ABA19906 Human ner

26 18.6 74.4 2585 5 ABA18948 Human ner

27 18.6 74.4 2585 5 ABA20196 Human ner

C 28 18.6 74.4 10279 6 ABL33591 Human imm

C 29 18.6 74.4 10279 6 ABL32277 Chemical

C 30 18.6 74.4 10279 6 RAD22328 Chemical

31 18.6 74.4 10311 4 AAK84424 Human imm

32 18.6 74.4 10312 4 AAK84423 Human imm

C 33 18.6 74.4 73334 6 ABL34124 Human imm

C 34 18.6 74.4 73334 6 ABL92318 Chemical

C 35 18.4 73.6 10891 6 ABL32465 Human imm

C 36 18.2 72.8 581 5 AAS33559 Human cDN

C 37 18.2 72.8 891 5 AAS34560 Human DNA

38 18.2 72.8 1486 3 AAZ97081 Human sec

39 18.2 72.8 1486 8 ACH66710 Novel hum

40 18.2 72.8 2121 6 AAL43413 A thalian

C 41 18.2 72.8 10467 6 ABL49301 Human pol

C 42 18.2 72.8 10872 4 AAL03182 Human rep

C 43 18.2 72.8 53905 7 ACF30939 Rice cult

C 44 18.2 72.8 76363 7 ACF30938 Rice cult

45 18.2 72.8 349881 9 ADC86642 Human GPC

ALIGNMENTS

RESULT 1
ACD28591
ID ACD28591 standard; DNA; 25 BP.

XX ACD28591;

DT 10-OCT-2003 (first entry)

DE E. coli ArsR binding oligonucleotide CHROMS1B.

XX ArsR; arsenic resistance operon; biosensor; ss; arsenic.

OS Escherichia coli.

PN US2003096275-A1.

PD 22-MAY-2003.

PF 15-AUG-2002; 2002US-00222952.

PR 20-AUG-2001; 2001US-0313714P.

XX (LAIN/) LAING L G.

PA Laing LG;

XX WPI; 2003-576876/54.

PT New system comprising isolated protein and nucleic acid, and a detection system that indicates a change in binding of the protein to the nucleic acid in the presence of the analyte, useful for detecting the presence of an analyte in a sample.

PS Claim 35; Page 15; 36pp; English.

XX The invention relates to a new system (biosensor) for detecting the presence of analyte in a sample comprising: (1) an isolated protein that specifically binds the analyte; (2) an isolated nucleic acid containing a specific binding sequence that is bound specifically by the protein; and (3) a detection system that indicates a change in binding of the protein to the nucleic acid in the presence of the analyte. Also included are a biosensor device for detecting the presence of an analyte in a sample, detecting the presence of an analyte in a sample, and an ArsR (encoded by part of the arsenic resistance operon of E. coli) protein comprising an amino acid sequence that is at least 90% identical to amino acids 1-97 of the ArsR sequence appearing as ABU63440 binding to a nucleic acid

CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the bottom strand of an oligonucleotide which binds
 CC to a chromosomally expressed Arsr protein and is used in the biosensor of
 CC the invention

CC SQ Sequence 25 BP; 13 A; 3 C; 2 G; 7 T; 0 U; 0 Other;

Query Match 100.0%; Score 25; DB 8; Length 25;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TAAGTCAAAAACATATATGACTTAA 25
 |||||
 Db 1 TAAGTCAAAAACATATATGACTTAA 25
 |||||

RESULT 2

ACD28590/c
 ID ACD28590 standard; DNA; 25 BP.

XX AC ACD28590;
 XX DT 10-OCT-2003 (first entry)
 XX DE E. coli Arsr binding oligonucleotide CHROMS17.
 XX KW Arsr; arsenic resistance operon; biosensor; ss; arsenic.
 XX OS Escherichia coli.
 XX PN US2003096275-A1.
 XX FD 22-MAY-2003.

XX PF 15-AUG-2002; 2002US-00222952.

XX PR 20-AUG-2001; 2001US-0313714P.

XX PA (LAIN/) LAING L G.

XX PI Laing LG;

XX DR WPI; 2003-576876/54.

XX PT New system comprising isolated protein and nucleic acid, and a detection
 PT system that indicates a change in binding of the protein to the nucleic
 PT acid in the presence of the analyte, useful for detecting the presence of
 PT analyte in a sample.

PS Claim 35; Page 15; 36pp; English.

XX The invention relates to a new system (biosensor) for detecting the
 CC presence of analyte in a sample comprising: (1) an isolated protein that
 CC specifically binds the analyte; (2) an isolated nucleic acid containing a
 CC specific binding sequence that is bound specifically by the protein; and
 CC (3) a detection system that indicates a change in binding of the protein
 CC to the nucleic acid in the presence of the analyte. Also included are a
 CC biosensor device for detecting the presence of an analyte in a sample,
 CC detecting the presence of an analyte in a sample, and an Arsr (encoded by
 CC part of the arsenic resistance operon of E. coli) protein comprising an
 CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
 CC the Arsr sequence appearing as ABU63440 binding to a nucleic acid
 CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the top strand of an oligonucleotide which binds to a
 CC chromosomally expressed Arsr protein and is used in the biosensor of the
 CC invention

XX SQ Sequence 25 BP; 7 A; 2 C; 3 G; 13 T; 0 U; 0 Other;

Query Match 100.0%; Score 25; DB 8; Length 25;
 Best Local Similarity 100.0%; Pred. No. 1.6;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TAAGTCAAAAACATATATGACTTAA 25
 |||||
 Db 25 TAAGTCAAAAACATATATGACTTAA 1
 |||||

RESULT 3

ACD28587
 ID ACD28587 standard; DNA; 40 BP.

XX AC ACD28587;
 XX DT 10-OCT-2003 (first entry)
 XX DE E. coli Arsr binding oligonucleotide CHROML13.
 XX KW Arsr; arsenic resistance operon; biosensor; ss; arsenic.

XX OS Escherichia coli.

XX PN US2003096275-A1.

XX PD 22-MAY-2003.

XX PF 15-AUG-2002; 2002US-00222952.

XX PR 20-AUG-2001; 2001US-0313714P.

XX PA (LAIN/) LAING L G.

XX PI Laing LG;

XX DR WPI; 2003-576876/54.

XX PT New system comprising isolated protein and nucleic acid, and a detection
 PT system that indicates a change in binding of the protein to the nucleic
 PT acid in the presence of the analyte, useful for detecting the presence of
 PT analyte in a sample.

PS Claim 35; Page 15; 36pp; English.

XX The invention relates to a new system (biosensor) for detecting the
 CC presence of analyte in a sample comprising: (1) an isolated protein that
 CC specifically binds the analyte; (2) an isolated nucleic acid containing a
 CC specific binding sequence that is bound specifically by the protein; and
 CC (3) a detection system that indicates a change in binding of the protein
 CC to the nucleic acid in the presence of the analyte. Also included are a
 CC biosensor device for detecting the presence of an analyte in a sample,
 CC detecting the presence of an analyte in a sample, and an Arsr (encoded by
 CC part of the arsenic resistance operon of E. coli) protein comprising an
 CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
 CC the Arsr sequence appearing as ABU63440 binding to a nucleic acid
 CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the bottom strand of an oligonucleotide which binds
 CC to a chromosomally expressed Arsr protein and is used in the biosensor of
 CC the invention

XX SQ Sequence 40 BP; 17 A; 5 C; 7 G; 11 T; 0 U; 0 Other;

Query Match 100.0%; Score 25; DB 8; Length 40;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TAAGTCAAAAACATATATGACTTAA 25
 |||||
 Db 1 TAAGTCAAAAACATATATGACTTAA 25
 |||||

RESULT 4

ACD28586/c
 ID ACD28586 standard; DNA; 42 BP.

```

XX AC 282586;
XX 10-OCT-2003 (first entry)
XX E. coli ArsR binding oligonucleotide CHROMALIT.
XX ArsR; arsenic resistance operon; biosensor; ss; arsenic.
XX Escherichia coli.
XX US2003096275-A1.
XX 22-MAY-2003.
XX 15-AUG-2002; 2002US-00222952.
XX 20-AUG-2001; 2001US-0313714P.
XX (LAIN/) LAING L G.
XX Laing LG;
XX WPI; 2003-576876/54.
XX New system comprising isolated protein and nucleic acid, and a detection
XX system that indicates a change in binding of the protein to the nucleic
XX acid in the presence of the analyte, useful for detecting the presence of
XX analyte in a sample.
XX Claim 35; Page 15; 36pp; English.
XX The invention relates to a new system (biosensor) for detecting the
XX presence of analyte in a sample comprising: (1) an isolated protein that
XX specifically binds the analyte; (2) an isolated nucleic acid containing a
XX specific binding sequence that is bound specifically by the protein; and
XX (3) a detection system that indicates a change in binding of the protein
XX to the nucleic acid in the presence of the analyte. Also included are a
XX biosensor device for detecting the presence of an analyte in a sample,
XX detecting the presence of an analyte in a sample, and an ArsR (encoded by
XX part of the arsenic resistance operon of E. coli) protein comprising an
XX amino acid sequence that is at least 90% identical to amino acids 1-97 of
XX the ArsR sequence appearing as AB03440 binding to a nucleic acid
XX sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
XX system is useful for detecting the presence of analyte in a sample. The
XX present sequence is the top strand of an oligonucleotide which binds to a
XX chromosomally expressed ArsR protein and is used in the biosensor of the
XX invention
XX Sequence 42 BP; 11 A; 8 C; 5 G; 18 T; 0 U; 0 Other;
Query Match 100.0%; Score 25; DB 8; Length 42;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAAGTCAAAAACATATATGACTTAA 25
Db 42 TAAGTCAAAAACATATATGACTTAA 18
RESULT 5
ACAL5469
ID ACAL5469 standard; DNA; 401 BP.
XX ACAL5469;
XX ACAL5469;
XX 27-OCT-2003 (revised)
XX 19-JUN-2003 (first entry)
XX Prokaryotic essential gene antisense oligonucleotide #3339.
XX Antisense; ss; prokaryotic essential gene; cell proliferation;
XX drug design.
XX Archaea.
XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen XL, Zvekind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX Claim 1; SEQ ID NO 3339; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway; (8)
XX required for proliferation, or that inhibits cellular proliferation; (9)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is one of the 6213
XX antisense sequences of the invention. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to
XX standardise OS field)
XX Sequence 401 BP; 113 A; 68 C; 101 G; 119 T; 0 U; 0 Other;
Query Match 100.0%; Score 25; DB 7; Length 401;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAAGTCAAAAACATATATGACTTAA 25
Db 104 TAAGTCAAAAACATATATGACTTAA 128
RESULT 6

```

```

AAS82555/c
ID AAS82555 standard; cDNA; 2182 BP.
XX AC
XX DT
XX AC AAS82555;
XX DT
XX DT 13-FEB-2002 (first entry)
XX DE
XX DE DNA encoding novel human diagnostic protein #18359.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS
XX OS Homo sapiens.
XX PN
XX PN WO200175067-A2.
XX PD
XX PD 11-OCT-2001.
XX PF
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA
XX PA (HYSE-) HYSEQ INC.
XX PI
XX PI Drmanac RT, Liu C, Tang YT;
XX DR
XX DR WPI; 2001-639362/73.
XX DR P-FSDB; ASG18368.
XX PT
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS
XX PS Claim 1; SEQ ID NO 18359; 103pp; English.
XX CC
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (III) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
XX CC coding sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pat_sequences
XX SQ
XX SQ Sequence 2182 BP; 516 A; 569 C; 582 G; 515 T; 0 U; 0 Other;
XX Query Match 100.0%; Score 25; DB 5; Length 2182;
XX Best Local Similarity 100.0%; Pred. No. 1.7;
XX Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAAGTCAAAAACATATATGACTTAA 25
DB 528 TAAGTCAAAAACATATATGACTTAA 504
RESULT 7
ABZ15842
ID ABZ15842 standard; DNA; 2000 BP.
XX AC
XX DT
XX AC ABZ15842;
XX DT
XX DT 21-JAN-2003 (first entry)
XX DE
XX DE Arabidopsis thaliana stress regulated gene SEQ ID NO 3647.
XX KW
XX KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX OS
XX OS Arabidopsis thaliana.
XX PN
XX PN WO200216655-A2.
XX PD
XX PD 28-FEB-2002.
XX PF
XX PF 24-AUG-2001; 2001WO-US026685.
XX PR
XX PR 24-AUG-2000; 2000US-0227866P.
XX PR 26-JAN-2001; 2001US-0264647P.
XX PR 22-JUN-2001; 2001US-0300111P.
XX PA
XX PA (SCRI ) SCRIPPS RES INST.
XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX PI
XX PI Harper JF, Kreps J, Wang X, Zhu T;
XX DR
XX DR WPI; 2002-304127/34.
XX PT
XX PT Identifying a stress condition to which a plant cell has been exposed and
XX PT producing plants with increased tolerance to these abiotic stresses.
XX PS
XX PS Claim 14; SEQ ID NO 3647; 577pp + Sequence Listing; English.
XX CC
XX CC The invention relates to identifying a stress condition to which a plant
XX CC cell has been exposed, comprising: (a) contacting nucleic acid
XX CC representative of expressed polynucleotides in the plant cell with an
XX CC array or probes representative of the plant cell genome; and (b)
XX CC detecting a profile of expressed polynucleotides in the plant cell
XX CC characteristic of a stress response. The method is useful in the
XX CC production of transgenic plants, cells and seeds and in producing plants
XX CC with increased tolerance to abiotic stress. The present sequence is that
XX CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
XX CC in methods of the invention. Note: The sequence data for this patent is
XX CC not represented in the printed specification but is based on sequence
XX CC information supplied to Derwent by the European Patent Office
XX SQ
XX SQ Sequence 2000 BP; 714 A; 318 C; 295 G; 673 T; 0 U; 0 Other;
XX Query Match 80.8%; Score 20.2; DB 6; Length 2000;
XX Best Local Similarity 88.0%; Pred. No. 1.2e+02;
XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TAAGTCAAAAACATATATGACTTAA 25
DB 951 TAATCATTAACATATATGACTTAA 975
RESULT 8
ABL33397/c
ID ABL33397 standard; DNA; 5641 BP.
XX AC
XX AC ABL33397;
XX DT
XX DT 26-MAR-2002 (first entry)
XX DE
XX DE Human immune system associated gene SEQ ID NO: 1370.
XX KW
XX KW Human; immune system disease; cytosine methylation; antiasthmatic;
XX KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;
XX KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
```

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 XX ds.
 XX Homo sapiens.
 XX WO200200928-A2.
 XX 03-JAN-2002.
 XX 02-JUL-2001; 2001WO-EP007537.
 XX 30-JUN-2000; 2000DE-01032529.
 XX 01-SEP-2000; 2000DE-01043826.
 XX (EPIG-) EPIGENOMICS AG.
 XX Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-130909/17.
 XX Nucleic acid comprising fragment of chemically modified gene, useful for
 XX diagnosis and treatment of diseases associated with abnormal cytosine
 XX methylation.
 XX Claim 1; SEQ ID NO 1370; 32pp + Sequence Listing; German.
 XX The present invention provides a number of human immune system associated
 XX genes which are modified by the methylation of cytosines. The sequences
 XX can be used in the diagnosis and treatment of immune system disorders,
 XX including eye diseases such as retinopathy, neovascular glaucoma and
 XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 XX diseases. The present sequence is a gene of the invention
 XX Sequence 5641 BP; 2137 A; 34 C; 1132 G; 2338 T; 0 U; 0 Other;
 XX
 Query Match 80.8%; Score 20.2; DB 6; Length 5641;
 Best Local Similarity 88.0%; Pred. No. 1.3e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 XX
 QY 1 TAAGTCAAAAACATATATGACTTAA 25
 DB 1536 TAATTCAAAAAATATATGACTTAA 1512
 XX
 RESULT 9
 AAD58282
 ID AAD58282 standard; DNA; 62782 BP.
 XX
 AC AAD58282;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Human tumour suppressor gene, Lmt intron 3 DNA.
 XX
 KW Tumour suppressor gene; Lmt; cancer; therapy; cytostatic; human; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO2003066869-A1.
 XX
 PD 14-AUG-2003.
 XX
 PP 07-FEB-2003; 2003WO-AU0000126.
 XX
 PR 07-FEB-2002; 2002AU-00000371.
 XX
 PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 XX
 PI Cook WD, Mccaw BJ;
 XX
 DR WPI; 2003-646311/61.
 XX
 PT New nucleic acid molecule, useful for screening a subject for the
 XX presence of an aberration in a gene encoding an LMT.
 XX
 PS Claim 10; Page 314-333; 373pp; English.
 XX
 DE The invention relates to novel tumour suppressor gene, referred to as
 XX Lmt. The invention also relates to the field of cancer therapy and cancer
 XX diagnostics. The nucleic acid molecule is useful for screening a subject
 XX for the presence of an aberration in a gene encoding an LMT. The present
 XX sequence is human Lmt intron 3 DNA
 XX
 SQ Sequence 62782 BP; 15441 A; 9614 C; 9374 G; 14439 T; 0 U; 13914 Other;
 XX
 Query Match 79.2%; Score 19.8; DB 8; Length 62782;
 Best Local Similarity 91.3%; Pred. No. 1.9e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 XX
 QY 2 AAGTCAAAACATATATGACTTAA 24
 DB 28437 ATGTCATAAACATATATGACTTAA 28459
 XX
 RESULT 10
 AAD58281
 ID AAD58281 standard; DNA; 62782 BP.
 XX
 AC AAD58281;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Human tumour suppressor gene, Lmt intron 2 DNA.
 XX
 KW Tumour suppressor gene; Lmt; cancer; therapy; cytostatic; human; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO2003066869-A1.
 XX
 PD 14-AUG-2003.
 XX
 PP 07-FEB-2003; 2003WO-AU0000126.
 XX
 PR 07-FEB-2002; 2002AU-00000371.
 XX
 PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 XX
 PI Cook WD, Mccaw BJ;
 XX
 DR WPI; 2003-646311/61.
 XX
 PT New nucleic acid molecule, useful for screening a subject for the
 XX presence of an aberration in a gene encoding an LMT.
 XX
 PS Claim 10; Page 314-333; 373pp; English.
 XX
 DE The invention relates to novel tumour suppressor gene, referred to as
 XX Lmt. The invention also relates to the field of cancer therapy and cancer
 XX diagnostics. The nucleic acid molecule is useful for screening a subject
 XX for the presence of an aberration in a gene encoding an LMT. The present
 XX sequence is human Lmt intron 2 DNA
 XX
 SQ Sequence 62782 BP; 15441 A; 9614 C; 9374 G; 14439 T; 0 U; 13914 Other;
 XX
 Query Match 79.2%; Score 19.8; DB 8; Length 62782;
 Best Local Similarity 91.3%; Pred. No. 1.9e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 XX
 QY 2 AAGTCAAAACATATATGACTTAA 24
 DB 28437 ATGTCATAAACATATATGACTTAA 28459
 XX

XX New nucleic acid molecule, useful for screening a subject for the
 PT presence of an aberration in a gene encoding an LMT.
 XX
 XX Claim 10; Page 333-358; 373pp; English.
 XX
 CC The invention relates to novel tumour suppressor gene, referred to as
 CC Lmt. The invention also relates to the field of cancer therapy and cancer
 CC diagnostics. The nucleic acid molecule is useful for screening a subject
 CC for the presence of an aberration in a gene encoding an LMT. The present
 CC sequence is human Lmt intron 3 DNA
 XX
 SQ Sequence 62782 BP; 15441 A; 9614 C; 9374 G; 14439 T; 0 U; 13914 Other;
 XX
 Query Match 79.2%; Score 19.8; DB 8; Length 62782;
 Best Local Similarity 91.3%; Pred. No. 1.9e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 XX
 QY 2 AAGTCAAAACATATATGACTTAA 24
 DB 28437 ATGTCATAAACATATATGACTTAA 28459
 XX
 RESULT 10
 AAD58281
 ID AAD58281 standard; DNA; 62782 BP.
 XX
 AC AAD58281;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Human tumour suppressor gene, Lmt intron 2 DNA.
 XX
 KW Tumour suppressor gene; Lmt; cancer; therapy; cytostatic; human; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO2003066869-A1.
 XX
 PD 14-AUG-2003.
 XX
 PP 07-FEB-2003; 2003WO-AU0000126.
 XX
 PR 07-FEB-2002; 2002AU-00000371.
 XX
 PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 XX
 PI Cook WD, Mccaw BJ;
 XX
 DR WPI; 2003-646311/61.
 XX
 PT New nucleic acid molecule, useful for screening a subject for the
 XX presence of an aberration in a gene encoding an LMT.
 XX
 PS Claim 10; Page 314-333; 373pp; English.
 XX
 DE The invention relates to novel tumour suppressor gene, referred to as
 XX Lmt. The invention also relates to the field of cancer therapy and cancer
 XX diagnostics. The nucleic acid molecule is useful for screening a subject
 XX for the presence of an aberration in a gene encoding an LMT. The present
 XX sequence is human Lmt intron 2 DNA
 XX
 SQ Sequence 62782 BP; 15441 A; 9614 C; 9374 G; 14439 T; 0 U; 13914 Other;
 XX
 Query Match 79.2%; Score 19.8; DB 8; Length 62782;
 Best Local Similarity 91.3%; Pred. No. 1.9e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 XX
 QY 2 AAGTCAAAACATATATGACTTAA 24
 DB 28437 ATGTCATAAACATATATGACTTAA 28459
 XX

```
RESULT 11
ADE11169_3
Continuation (4 of 4) of ADE11169 from base 300001 (Human transporter protein encoding g
WP Sequence split into 4 fragments LOCUS ADE11169 Accession ADE11169
WP Fragment Name Begin End
WP ADE11169_0 1 110000
WP ADE11169_1 100001 210000
WP ADE11169_2 200001 310000
WP ADE11169_3 300001 394191

Query Match 79.2%; Score 19.8; DB 9; Length 94191;
Best Local Similarity 91.3%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAAGTCAAAACATATATGACTT 23
|||||
DB 35185 TAAGTCAAAACATATATGACTT 35207

RESULT 12
AAD58279
ID AAD58279 standard; DNA; 226475 BP.
AC AAD58279;
XX
XX 20-NOV-2003 (first entry)
DT Human tumour suppressor gene, lmt reverse complement DNA.
DE Tumour suppressor gene; lmt; cancer; therapy; cytostatic; human; ds.
XX Homo sapiens.
XX WO200306869-A1.
XX 14-AUG-2003.
XX
XX 07-FEB-2003; 2003WO-AU000126.
XX
XX 07-FEB-2002; 2002AU-00000371.
XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
XX Cook WD, Mccaw BJ;
XX WPI; 2003-646311/61.
XX
XX New nucleic acid molecule, useful for screening a subject for the
XX presence of an aberration in a gene encoding an LMT.
XX Claim 10; Page 233-299; 373pp; English.
XX
XX The invention relates to novel tumour suppressor gene, referred to as
XX Lmt. The invention also relates to the field of cancer therapy and cancer
XX diagnostics. The nucleic acid molecule is useful for screening a subject
XX for the presence of an aberration in a gene encoding an LMT. The present
XX sequence is human lmt reverse complement DNA
XX
XX Sequence 226475 BP; 61024 A; 41761 C; 40916 G; 57494 T; 0 U; 25280 Other;

Query Match 79.2%; Score 19.8; DB 8; Length 226475;
Best Local Similarity 91.3%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAGTCAAAACATATATGACTTA 24
|||||
DB 140928 ATGTCATAAACATATATGACTTA 140950

RESULT 13
AAC36413/c
ID AAC36413 standard; DNA; 1293 BP.
XX
```

AAC36413;
17-OCT-2000 (first entry)
Arabidopsis thaliana DNA fragment SEQ ID NO: 13710.
Hybridisation assay; Genetic mapping; gene expression control;
protein identification; signal transduction pathway; metabolic pathway;
promoter; termination sequence; ss.
Arabidopsis thaliana.
EP1033405-A2.
06-SEP-2000.
25-FEB-2000; 2000EP-00301439.
25-FEB-1999; 99US-0121825P.
05-MAR-1999; 99US-0123180P.
09-MAR-1999; 99US-0123548P.
23-MAR-1999; 99US-0125788P.
25-MAR-1999; 99US-0126264P.
23-MAR-1999; 99US-0126785P.
01-APR-1999; 99US-0127462P.
06-APR-1999; 99US-0128234P.
08-APR-1999; 99US-0128714P.
16-APR-1999; 99US-0129845P.
19-APR-1999; 99US-0130077P.
21-APR-1999; 99US-0130449P.
23-APR-1999; 99US-0130510P.
23-APR-1999; 99US-0130891P.
28-APR-1999; 99US-0131449P.
30-APR-1999; 99US-0132048P.
04-MAY-1999; 99US-0132407P.
05-MAY-1999; 99US-0132484P.
06-MAY-1999; 99US-0132485P.
06-MAY-1999; 99US-0132486P.
07-MAY-1999; 99US-0132487P.
11-MAY-1999; 99US-0132863P.
14-MAY-1999; 99US-0134256P.
14-MAY-1999; 99US-0134218P.
14-MAY-1999; 99US-0134219P.
14-MAY-1999; 99US-0134221P.
18-MAY-1999; 99US-0134370P.
18-MAY-1999; 99US-0134768P.
19-MAY-1999; 99US-0134941P.
20-MAY-1999; 99US-0135124P.
21-MAY-1999; 99US-0135353P.
24-MAY-1999; 99US-0135629P.
25-MAY-1999; 99US-0136021P.
27-MAY-1999; 99US-0136392P.
27-MAY-1999; 99US-0136782P.
01-JUN-1999; 99US-0137222P.
03-JUN-1999; 99US-0137528P.
04-JUN-1999; 99US-0137502P.
07-JUN-1999; 99US-0137724P.
08-JUN-1999; 99US-0138094P.
10-JUN-1999; 99US-0138540P.
10-JUN-1999; 99US-0138847P.
14-JUN-1999; 99US-0139119P.
16-JUN-1999; 99US-0139452P.
16-JUN-1999; 99US-0139453P.
17-JUN-1999; 99US-0139492P.
18-JUN-1999; 99US-0139454P.
18-JUN-1999; 99US-0139455P.
18-JUN-1999; 99US-0139456P.
18-JUN-1999; 99US-0139457P.
18-JUN-1999; 99US-0139458P.
18-JUN-1999; 99US-0139459P.
18-JUN-1999; 99US-0139460P.
18-JUN-1999; 99US-0139461P.
18-JUN-1999; 99US-0139462P.

```
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140351P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140693P.
PR 24-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145911P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0145386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149922P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
```

```
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160788P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 22-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
```

Query Match 76.8%; Score 19.2; DB:3; Length 1293;

Best Local Similarity 87.5%; Pred. No. 3e+02; Mismatches 0; Gaps 0;

Matches 21; Conservative 0; Indels 3; Indels 0; Gaps 0;

Qy 2 AAGTCAAAACATATATGACTTAA 25

Db 387 AAGTCATCAACATATAAGACTTAA 364

RESULT 14

ABA92787_2

Continuation (3 of 7) of ABA92787 from base 200001 (Buchnera sp. genomic DNA SEQ ID NO: 1)

WP	Sequence split into 7 fragments	LOCUS ABA92787 Accession ABA92787
WP	Fragment Name	Begin
WP	ABA92787_0	1
WP	ABA92787_1	100001
WP	ABA92787_2	200001
WP	ABA92787_3	300001
WP	ABA92787_4	400001
WP	ABA92787_5	500001

WP ABA92787_6 600001 640681
Query Match 75.2%; Score 18.8; DB 6; Length 110000;
Best Local Similarity 90.9%; Pred No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAGTCAAAACATATATGACTT 23
|||||||
Db 23317 AAGTCAAAACATATATGATT 23338

RESULT 15
ABAI3700
ID ABAI3700 standard; cDNA; 561 BP.
XX
AC ABAI3700;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 2707.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antispasmodic; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
XX WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US0001334.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214866P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241828P.
PR 01-NOV-2000; 2000US-0242221P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.

PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 01-DEC-2000; 2000US-0251160P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI; 2001-541565/60.
 DR P-PSDB; ABB17374.
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system cancers
 PT and metastases.
 XX
 PS Claim 1; SEQ ID NO 2707; 1701pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
 CC (ABA14678-ABB18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 561 BP; 217 A; 96 C; 84 G; 155 T; 0 U; 9 Other;
 Query Match 74.4%; Score 18.6; DB 5; Length 561;
 Best Local Similarity 84.0%; Pred. No. 5.1e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 TAAGTCAAAAACATATATGACTTAA 25
 DB 383 TGAGTCAAAAACAAACATGCTTAA 407

Search completed: May 26, 2004, 17:50:16
 Job time : 178.526 secs

RESULT 2
US-08-913-014A-9
: Sequence 9, Application US/08913014A
: Patent No. 6235878
: GENERAL INFORMATION:
: APPLICANT: Nishi, Kazunori
: APPLICANT: Hikichi, Yukiko
: APPLICANT: Shintani, Yasushi
: TITLE OF INVENTION: NOVEL FAS LIGAND-LIKE PROTEIN, ITS
: TITLE OF INVENTION: PRODUCTION AND USE
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David G. Conlin, Esq.
: ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & USHMAN, LLP
: STREET: 130 Water Street
: CITY: Boston,
: STATE: MA
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS

```
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913.014A
; FILING DATE: 04-SEP-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP97/02480
; FILING DATE: July 17, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: David G. Conlin
; REGISTRATION NUMBER: 27,026
; REFERENCE/DOCKET NUMBER: 342/47694
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9058
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Linear
; MOLECULE TYPE: genomic DNA
; US-08-913-014A-9

Query Match 66.7%; Score 20; DB 3; Length 9058;
Best Local Similarity 82.1%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TAATCATATGCGTTTTGGTTATGTT 29
Db 8577 TAATCATTTGCTTTTGGTTATTTGTT 8604

RESULT 3
US-09-653-285-9
; Sequence 9, Application US/09653285
; Patent No. 6590090
; GENERAL INFORMATION:
; APPLICANT: Nishi, Kazunori
; Hixichi, Yukiko
; TITLE OF INVENTION: NOVEL PAS LIGAND-LIKE PROTEIN, ITS
; PRODUCTION AND USE
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David G. Conlin, Esq.
; STREET: 130 Water Street
; CITY: Boston,
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/653,285
; FILING DATE: 31-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/913,014
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: David G. Conlin
; REGISTRATION NUMBER: 27,026
; REFERENCE/DOCKET NUMBER: 342/47694
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9058
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Linear
; MOLECULE TYPE: genomic DNA
; US-09-653-285-9

Query Match 66.7%; Score 20; DB 4; Length 9058;
Best Local Similarity 82.1%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TAATCATATGCGTTTTGGTTATGTT 29
Db 8577 TAATCATTTGCTTTTGGTTATTTGTT 8604

RESULT 4
US-09-328-352-3389
; Sequence 3389, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3389
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-3389

Query Match 64.7%; Score 19.4; DB 4; Length 2481;
Best Local Similarity 79.3%; Pred. No. 34;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTTGGTTATGTT 29
Db 1913 TAAATGATATGCTCTTGGTTTATGGTT 1941

RESULT 5
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHIRO
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
; US-09-790-988-1

Query Match 64.7%; Score 19.4; DB 4; Length 640681;
Best Local Similarity 79.3%; Pred. No. 66;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

Qy 1 TTAATCATATGCGTTTGGTTATGTTG 29
|||||
Db 516370 TTAATAATCTAAGTTTGGTTAAATGTT 516342

RESULT 6

US-09-422-978-2280

; Sequence 2280, Application US/09422978

; Patent No. 6537751

; GENERAL INFORMATION:

; APPLICANT: Cohen, Daniel

; APPLICANT: Blumenfeld, Marta

; APPLICANT: Chumakov, Iliya

; TITLE OF INVENTION: Biallelic markers for use in constructing a high density...

; FILE REFERENCE: GENSET.020CP1

; CURRENT APPLICATION NUMBER: US/09/422,978

; EARLIER FILING DATE: 1999-10-20

; EARLIER APPLICATION NUMBER: US 09/298,850

; EARLIER FILING DATE: 1999-04-21

; EARLIER APPLICATION NUMBER: US 60/109,732

; EARLIER FILING DATE: 1998-11-23

; EARLIER APPLICATION NUMBER: US 60/082,614

; EARLIER FILING DATE: 1998-04-21

; NUMBER OF SEQ ID NOS: 11796

; SEQ ID NO 2280

; LENGTH: 47

; TYPE: DNA

; ORGANISM: Homo Sapiens

; FEATURE:

; NAME/KEY: allele

; LOCATION: 24

; OTHER INFORMATION: 99-10146-202 : polymorphic base T or A

US-09-422-978-2280

Query Match 64.0%; Score 19.2; DB 4; Length 47;

Best Local Similarity 80.8%; Pred. No. 25;

Matches 21; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 3 AATCATATGCGTTTGGTTATGTTG 28

|||||

Db 22 AATATATGCAATTTGTTATGTTT 47

RESULT 7

US-09-349-740A-16/c

; Sequence 16, Application US/09349740A

; Patent No. 6476297

; GENERAL INFORMATION:

; APPLICANT: Mascarenhas, Joseph P.

; APPLICANT: He, Caiping

; TITLE OF INVENTION: Meiosis-Related Gene and Promoter

; FILE REFERENCE: 0794.009A

; CURRENT APPLICATION NUMBER: US/09/349,740A

; CURRENT FILING DATE: 1999-07-08

; PRIOR APPLICATION NUMBER: 60/092,277

; PRIOR FILING DATE: 1998-07-10

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: WordPerfect 8.0

; SEQ ID NO 16

; LENGTH: 2911

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

; FEATURE:

; NAME/KEY: promoter

; LOCATION: (1)...(2911)

; NAME/KEY: allele

; LOCATION: (450)

; OTHER INFORMATION: n may be C,G,T or A

; NAME/KEY: allele

; LOCATION: (546)

; OTHER INFORMATION: n may be C,G,T or A

; NAME/KEY: allele

; LOCATION: (737)

; OTHER INFORMATION: n may be C,G,T or A

US-09-349-740A-16

Query Match 62.7%; Score 18.8; DB 4; Length 2911;

Best Local Similarity 76.7%; Pred. No. 61;

Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 TTAATCATATGCGTTTGGTTATGTTG 30

|||||

Db 2051 TTCAACATATCGGAATTAGGTTCTTTGTTG 2022

RESULT 8

US-09-349-740A-1/c

; Sequence 1, Application US/09349740A

; Patent No. 6476297

; GENERAL INFORMATION:

; APPLICANT: Mascarenhas, Joseph P.

; APPLICANT: He, Caiping

; TITLE OF INVENTION: Meiosis-Related Gene and Promoter

; FILE REFERENCE: 0794.009A

; CURRENT APPLICATION NUMBER: US/09/349,740A

; CURRENT FILING DATE: 1999-07-08

; PRIOR APPLICATION NUMBER: 60/092,277

; PRIOR FILING DATE: 1998-07-10

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: WordPerfect 8.0

; SEQ ID NO 1

; LENGTH: 4500

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

; FEATURE:

; NAME/KEY: allele

; LOCATION: (450)

; OTHER INFORMATION: n may be C,G,T or A

; NAME/KEY: allele

; LOCATION: (546)

; OTHER INFORMATION: n may be C,G,T or A

; NAME/KEY: allele

; LOCATION: (737)

; OTHER INFORMATION: n may be C,G,T or A

US-09-349-740A-1

Query Match 62.7%; Score 18.8; DB 4; Length 4500;

Best Local Similarity 76.7%; Pred. No. 64;

Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 TTAATCATATGCGTTTGGTTATGTTG 30

|||||

Db 2051 TTCAACATATCGGAATTAGGTTCTTTGTTG 2022

RESULT 9

US-09-596-002-41/c

; Sequence 41, Application US/09596002

; Patent No. 6632636

; GENERAL INFORMATION:

; APPLICANT: Lagace, Robert, E.

; APPLICANT: Patterson, Chandra

; APPLICANT: Berg, Kim, L.

; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME

; FILE REFERENCE: PM-0008-4 US

; CURRENT APPLICATION NUMBER: US/09/596,002

; CURRENT FILING DATE: 2000-06-16

; PRIOR APPLICATION NUMBER: 60/140,121

; PRIOR FILING DATE: 1999-06-18

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PERL Program

; SEQ ID NO 41

; LENGTH: 269223

; TYPE: DNA

; ORGANISM: Moraxella catarrhalis

; FEATURE:

NAME/KEY: misc feature
OTHER INFORMATION: Incyte template ID No. 6632636 41
PUBLICATION INFORMATION:
US-09-557-884-1

Query Match 62.7%; Score 18.8; DB 4; Length 269223;
Best Local Similarity 90.9%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTAATCATATCGCTTTTGGTT 22
Db 142085 TTAATCATATCGCTTTTGGCT 142064

RESULT 10

US-09-557-884-1/c
Sequence 1, Application US/09557884
Patent No. 6506581
GENERAL INFORMATION:

APPLICANT: Fleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884

FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995

ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-557-884-1

Query Match 62.7%; Score 18.8; DB 4; Length 1830121;
Best Local Similarity 90.9%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AATCATATCGCTTTTGGTTAT 24
Db 1238911 AACATATCGCTTTTGGTTAT 1238890

RESULT 11

US-09-643-990A-1/c
Sequence 1, Application US/09643990A
Patent No. 6528289
GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 62.7%; Score 18.8; DB 4; Length 1830121;
Best Local Similarity 90.9%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AATCATATCGCTTTTGGTTAT 24
Db 1238911 AACATATCGCTTTTGGTTAT 1238890

RESULT 12

US-09-227-357-119
Sequence 119, Application US/09227357
Patent No. 6342581
GENERAL INFORMATION:
APPLICANT: Fischer et al.
TITLE OF INVENTION: 123 Human Secreted Proteins
FILE REFERENCE: P2010P1
CURRENT APPLICATION NUMBER: US/09/227,357
CURRENT FILING DATE: 1999-01-08
EARLIER APPLICATION NUMBER: PCT/US98/13684
EARLIER FILING DATE: 1998-07-07
EARLIER APPLICATION NUMBER: 60/051,926
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,793
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,925
EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/051,929
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,803
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,732
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,931
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,932
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,916
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,930
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,918
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,920
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,733
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,795
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,919
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,928
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/055,722
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,723
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,948
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,949
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,953
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,950
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,947
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,964
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/056,360
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,684
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,984
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,954
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/058,785
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,664
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,660
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,661
EARLIER FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 672
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 119
LENGTH: 1442
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: {1377}
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: {1419}
OTHER INFORMATION: n equals a,t,g, or c
US-09-227-357-119

Query Match 62.0%; Score 18.6; DB 4; Length 1442;
Best Local Similarity 84.0%; Pred. No. 67;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 TAATCATATGCGTTTGGTTATGCT 26
|||||
DB 750 TAATCAATTTGTTATGCTGTTGCT 774
|||||

RESULT 13

US-09-107-532A-1957
Sequence 1957, Application US/09107532A
Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOMB THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 1957:

SEQUENCE CHARACTERISTICS:

LENGTH: 567 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...567
SEQUENCE DESCRIPTION: SEQ ID NO: 1957:
US-09-107-532A-1957

Query Match 61.3%; Score 18.4; DB 4; Length 567;
Best Local Similarity 78.6%; Pred. No. 72;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 AATCATATGCGTTTGGTTATGCTG 30
|||||
DB 93 AATCATATGCGATTGATATGTTG 120
|||||

RESULT 14

US-09-539-333D-178
; Sequence 178, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bouguetere, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: GENSET.047AUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: US 09/416,384
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent.pm
; SEQ ID NO 178
; LENGTH: 3001
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1501
; OTHER INFORMATION: 99-15682-318 : polymorphic base A or T
; FEATURE:
; NAME/KEY: misc.binding
; LOCATION: 1482..1500
; OTHER INFORMATION: 99-15682-318.mis1
; FEATURE:
; NAME/KEY: misc.binding
; LOCATION: 1502..1521
; OTHER INFORMATION: 99-15682-318.mis2, complement
; FEATURE:
; NAME/KEY: primer.bind
; LOCATION: 1184..1202
; OTHER INFORMATION: upstream amplification primer
; FEATURE:
; NAME/KEY: primer.bind
; LOCATION: 1665..1683
; OTHER INFORMATION: downstream amplification primer, complement
; FEATURE:
; NAME/KEY: misc.binding
; LOCATION: 1489..1513
; OTHER INFORMATION: 99-15682-318 probe
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1842_
; OTHER INFORMATION: n=a, g, c or t
; US-09-539-333D-178

Query Match 61.3%; Score 18.4; DB 4; Length 3001;
Best Local Similarity 78.6%; Pred. No. 89;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 TAATCATATGCGTTTGGTTATGTTT 29

Db 2843 TAATATATATGTTTATGTTATCTTT 2870
RESULT 15
US-10-204-708-39
; Sequence 39, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIRPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 39
; LENGTH: 19513
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)
; US-10-204-708-39

Query Match 61.3%; Score 18.4; DB 4; Length 19513;
Best Local Similarity 78.6%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 TAATCATATGCGTTTGGTTATGTTT 29
Db 18350 TTATATATGCGTTTGGTTATGTTT 18377

Search completed: May 26, 2004, 17:56:28
Job time : 54.8205 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 17:50:29 ; Search time 615.897 Seconds
(without alignments)
221.574 Million cell updates/sec

Title: US-10-676-299-3

Perfect score: 30
Sequence: 1 ttaacatcatgcgttttggtatgtgtg 30

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2960401 seqs, 2274450654 residues

Total number of hits satisfying chosen parameters: 5920802

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgm2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgm2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgm2_6/ptodata/2/pubpna/US05_NEW_PUB.seq.*
4: /cgm2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgm2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgm2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgm2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgm2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
9: /cgm2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
10: /cgm2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
11: /cgm2_6/ptodata/2/pubpna/US09D_PUBCOMB.seq.*
12: /cgm2_6/ptodata/2/pubpna/US09E_NEW_PUB.seq.*
13: /cgm2_6/ptodata/2/pubpna/US09F_NEW_PUB.seq.*
14: /cgm2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgm2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgm2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
17: /cgm2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
18: /cgm2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
19: /cgm2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	
1	30	100.0	30	15	US-10-232-952A-3	
C	2	28	93.3	28	15	US-10-232-952A-4
	3	23	76.7	23	15	US-10-232-952A-7
4	22	73.3	248436	13	US-10-087-192-2014	
5	21.6	72.0	6391	15	US-10-240-453-11	
C	6	21.6	72.0	79860	13	US-10-087-192-412
7	21.2	70.7	343	9	US-09-867-701-10855	
C	8	21	70.0	21	15	US-10-232-952A-8
C	9	21	70.0	741	13	US-10-027-632-166036
C	10	21	70.0	741	15	US-10-027-632-166036
11	20.4	68.0	6022	13	US-10-221-714A-119	
12	20.4	68.0	6022	13	US-10-239-676-79	
13	20.4	68.0	6022	15	US-10-311-455-1045	
14	20.4	68.0	6022	15	US-10-240-453-85	
					Sequence 3, Appli	
					Sequence 4, Appli	
					Sequence 7, Appli	
					Sequence 11, Appl	
					Sequence 14, Ap	
					Sequence 412, App	
					Sequence 10855, A	
					Sequence 8, Appli	
					Sequence 166036,	
					Sequence 119, App	
					Sequence 79, Appl	
					Sequence 1045, Ap	
					Sequence 85, Appl	

c 15 20 66.7 520 13 US-10-027-632-204483 Sequence 204483,
c 16 20 66.7 520 16 US-10-027-632-204483 Sequence 204483,
c 17 20 66.7 520 13 US-10-027-632-129002 Sequence 129002,
c 18 20 66.7 520 16 US-10-027-632-129002 Sequence 129002,
c 19 20 66.7 520 13 US-10-027-632-129002 Sequence 129002,
c 20 20 66.7 520 13 US-10-027-632-129002 Sequence 129002,
c 21 19.8 66.0 251364 15 US-10-175-523-58 Sequence 58, Appl
c 22 19.8 66.0 251364 15 US-10-175-523-61 Sequence 61, Appl
c 23 19.8 66.0 251364 15 US-10-175-523-79 Sequence 79, Appl
c 24 19.6 65.3 5823 15 US-10-240-453-256 Sequence 256, Appl
c 25 19.6 65.3 378361 10 US-09-901-136-3 Sequence 3, Appli
c 26 19.4 64.7 501 13 US-10-027-632-64934 Sequence 64934, A
c 27 19.4 64.7 501 13 US-10-027-632-296688 Sequence 296688,
c 28 19.4 64.7 501 16 US-10-027-632-64934 Sequence 64934, A
c 29 19.4 64.7 501 16 US-10-027-632-296688 Sequence 296688,
c 30 19.4 64.7 501 13 US-10-027-632-37667 Sequence 37667, A
c 31 19.4 64.7 501 13 US-10-027-632-37667 Sequence 37667, A
c 32 19.4 64.7 501 13 US-10-027-632-289306 Sequence 289306,
c 33 19.4 64.7 501 13 US-10-027-632-289306 Sequence 289306,
c 34 19.4 64.7 501 16 US-10-027-632-289306 Sequence 289306,
c 35 19.4 64.7 501 16 US-10-027-632-289306 Sequence 289306,
c 36 19.4 64.7 501 13 US-10-027-632-289306 Sequence 289306,
c 37 19.4 64.7 501 13 US-10-027-632-289306 Sequence 289306,
c 38 19.4 64.7 501 15 US-10-094-749-622 Sequence 142009,
c 39 19.4 64.7 501 15 US-10-311-455-319 Sequence 319, Appl
c 40 19.4 64.7 501 15 US-10-240-453-11 Sequence 11, Appl
c 41 19.4 64.7 501 15 US-10-240-453-11 Sequence 11, Appl
c 42 19.4 64.7 501 13 US-10-221-613-401 Sequence 401, Appl
c 43 19.4 64.7 501 15 US-10-311-455-1379 Sequence 1379, Ap
c 44 19.4 64.7 501 10 US-09-764-891-7733 Sequence 7733, Ap
c 45 19.4 64.7 501 15 US-10-239-676-105 Sequence 105, Appl
c 46 19.4 64.7 501 15 US-10-311-455-1315 Sequence 1315, Ap

ALIGNMENTS

RESULT 1
US-10-222-952A-3
; Sequence 3, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regensis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/1L443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PLASMIT biotinylated top, long oligo sequence; biotinylated
; OTHER INFORMATION: nucleotide at position 1
US-10-222-952A-3

Query Match 100.0%; Score 30; DB 15; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.052;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTAATCATATGCGTTTGTGTTATGTTG 30
Db 1 TTAATCATATGCGTTTGTGTTATGTTG 30

RESULT 2
US-10-222-952A-4/c
; Sequence 4, Application US/10222952A
; Publication No. US20030096275A1

GENERAL INFORMATION:
APPLICANT: Regensis
APPLICANT: Laing, Lance
TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
FILE REFERENCE: 4107/1L443-US1
CURRENT APPLICATION NUMBER: US/10/222,952A
CURRENT FILING DATE: 2002-08-15
PRIOR APPLICATION NUMBER: US 60/313,714
PRIOR FILING DATE: 2001-08-20
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 28
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PLAS1B bottom, long oligo sequence
US-10-222-952A-4

Query Match 93.3%; Score 28; DB 15; Length 28;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AATCATATGCGTTTGGTTATGTTG 30
|||||
Db 28 AATCATATGCGTTTGGTTATGTTG 1

RESULT 3
US-10-222-952A-7
Sequence 7, Application US/10222952A
Publication No. US20030096275A1
GENERAL INFORMATION:
APPLICANT: Regensis
APPLICANT: Laing, Lance
TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
FILE REFERENCE: 4107/1L443-US1
CURRENT APPLICATION NUMBER: US/10/222,952A
CURRENT FILING DATE: 2002-08-15
PRIOR APPLICATION NUMBER: US 60/313,714
PRIOR FILING DATE: 2001-08-20
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7
LENGTH: 23
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PLAS1T top, short biotinylated oligo sequence; biotinylated
OTHER INFORMATION: nucleotide at position 1
US-10-222-952A-7

Query Match 76.7%; Score 23; DB 15; Length 23;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTGGTTA 23
|||||
Db 1 TTAATCATATGCGTTTGGTTA 23

RESULT 4
US-10-087-192-2014
Sequence 2014, Application US/10087192
Publication No. US20020182586A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
CURRENT FILING DATE: 2002-03-01

PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2014
LENGTH: 248436
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(248436)
OTHER INFORMATION: n = A,T,C or G
US-10-087-192-2014

Query Match 73.3%; Score 22; DB 13; Length 248436;
Best Local Similarity 83.3%; Pred. No. 2.6e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTGGTTATGTTG 30
|||||
Db 76015 TTAATCATCTGTGTTTGGTTATTTGTTG 76044

RESULT 5
US-10-240-453-11
Sequence 11, Application US/10240453
Publication No. US20030148326A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
TITLE OF INVENTION: Transcription
TITLE OF INVENTION: By Means of Assessing the Methylation Status of Genes Associati
TITLE OF INVENTION: With DNA Transcription
FILE REFERENCE: 5013,1009
CURRENT APPLICATION NUMBER: US/10/240,453
CURRENT FILING DATE: 2002-10-02
PRIOR APPLICATION NUMBER: PCT/EP01/03973
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 350
SEQ ID NO 11
LENGTH: 6391
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-11

Query Match 72.0%; Score 21.6; DB 15; Length 6391;
Best Local Similarity 85.7%; Pred. No. 2e+02; 4; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0;

QY 2 TTAATCATATGCGTTTGGTTATGTTG 29
|||||
Db 1000 TTAATAATATGTTTGGATATGTTT 1027

RESULT 6
US-10-087-192-412/c
Sequence 412, Application US/10087192
Publication No. US20020182586A1
GENERAL INFORMATION:

; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 412
; LENGTH: 79860
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (1)...(79860)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-412

Query Match 72.0%; Score 21.6; DB 13; Length 79860;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AATCATATGCGTTTGGTTATGTTG 30
|||||
Db 65299 ACTCATATGCTTTTGGTATGAGTTG 65272

RESULT 7
US-09-867-701-10855
; Sequence 10855, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Hatlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10855
; LENGTH: 343
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-10855

Query Match 70.7%; Score 21.2; DB 9; Length 343;
Best Local Similarity 88.5%; Pred. No. 1.7e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTGGTTATGT 26
|||||
Db 78 TTAATCAGATGCGCTTTTAGTTATGT 103

RESULT 8
US-10-222-952A-8/c
; Sequence 8, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regenesis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/11443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15

; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: FLASS1B bottom, short oligo sequence
US-10-222-952A-8

Query Match 70.0%; Score 21; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AATCATATGCGTTTGGTTA 23
|||||
Db 21 AATCATATGCGTTTGGTTA 1

RESULT 9
US-10-027-632-166036/c
; Sequence 166036, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 166036
; LENGTH: 741
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-166036

Query Match 70.0%; Score 21; DB 13; Length 741;
Best Local Similarity 82.8%; Pred. No. 2.3e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TAATCATATGCGTTTGGTTATGTTG 30
|||||
Db 303 TAATCAGATGCTTTTGGTTGTTTG 275

RESULT 10
US-10-027-632-166036/c
; Sequence 166036, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632

```
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 166036
; LENGTH: 741
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-166036
```

```
Query Match 70.0%; Score 21; DB 16; Length 741;
Best Local Similarity 82.8%; Pred. No. 2.3e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
Oy 2 TAATCATATGCGTTTGTGTTGTTG 30
|||||
Db 303 TAATCATATGAGTTTGTGTTGTTG 275
```

```
RESULT 11
US-10-221-714A-119
; Sequence 119, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 119
; LENGTH: 6022
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-119
```

```
Query Match 68.0%; Score 20.4; DB 13; Length 6022;
Best Local Similarity 80.0%; Pred. No. 5.5e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
Oy 1 TTAATCATATGCGTTTGTGTTGTTG 30
|||||
Db 2072 TTAATCATATATTTTGTGTTGTTG 2101
```

```
RESULT 12
US-10-239-676-79
; Sequence 79, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 79
; LENGTH: 6022
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-79
```

```
Query Match 69.0%; Score 20.4; DB 15; Length 6022;
Best Local Similarity 80.0%; Pred. No. 5.5e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
Oy 1 TTAATCATATGCGTTTGTGTTGTTG 30
|||||
Db 2072 TTAATCATATATTTTGTGTTGTTG 2101
```

```
RESULT 13
US-10-311-455-1045
; Sequence 1045, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Dete
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1045
; LENGTH: 6022
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1045
```

```
Query Match 68.0%; Score 20.4; DB 15; Length 6022;
Best Local Similarity 80.0%; Pred. No. 5.5e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
OY 1 TTAATCATATGCGTTTGGTTATGTTG 30
Db 2072 TTAATGATATATTATTTTGGTTATGTAATG 2101

RESULT 14
US-10-240-453-85
; Sequence 85, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
; TITLE OF INVENTION: With DNA Transcription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240,453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 85
; LENGTH: 6022
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-85

Query Match 68.0%; Score 20.4; DB 15; Length 6022;
Best Local Similarity 80.0%; Pred. No. 5.5e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 TTAATCATATGCGTTTGGTTATGTTG 30
Db 2072 TTAATGATATATTATTTTGGTTATGTAATG 2101

RESULT 15
US-10-027-632-204483/C
; Sequence 204483, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
```

```
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 204483
; LENGTH: 520
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-204483

Query Match 66.7%; Score 20; DB 13; Length 520;
Best Local Similarity 82.1%; Pred. No. 5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 TAAATCATATGCGTTTGGTTATGTTG 29
Db 280 TATTCATATGCTTATTTGGTTTTCCT 253

Search completed: May 27, 2004, 06:16:13
Job time : 618.897 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 15:22:28 ; Search time 214.231 Seconds
(without alignments)

594.900 Million cell updates/sec

Title: US-10-676-299-3

Perfect score: 30

Sequence: 1 ttaatcatatgcgttttggtagtggtg 30

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Geneseqn1980a:*
- 2: Geneseqn1990a:*
- 3: Geneseqn2000a:*
- 4: Geneseqn2001a:*
- 5: Geneseqn2002a:*
- 6: Geneseqn2003a:*
- 7: Geneseqn2004a:*
- 8: Geneseqn2005a:*
- 9: Geneseqn2006a:*
- 10: Geneseqn2007a:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	30	8	ACD28584
C	28	93.3	28	8	ACD28585 E. coli A
	23	76.7	23	8	ACD28588 E. coli A
	21.6	72.0	6391	6	ABK28137 DNA trans
C	5	21.6	72.0	23532	3 AAH81455 N. mening
C	6	21.6	72.0	37668	3 AAH81490_14 Continuation (15 o
C	7	21.6	72.0	110000	3 AAH81490_13 Continuation (14 o
C	8	21.6	72.0	349980	3 AAF21610 Neisseria
9	21.2	70.7	343	6	ABU67877 Human ova
C	10	21	70.0	21	ACD28589 E. coli A
11	20.6	68.7	459	8	ADA29013 DNA encod
12	20.4	68.0	1137	4	AB116815 Drosophil
13	20.4	68.0	1543	4	AAF56298 IS2 eleme
14	20.4	68.0	6022	4	AAH45374 Chemical
15	20.4	68.0	6022	4	AAH46397 Tumour su
16	20.4	68.0	6022	6	AB133072 Human imm
17	20.4	68.0	6022	6	ABK28211 DNA trans
C	18	20	66.7	1000	4 AAF91383 N. mening
C	19	20	66.7	1000	6 ABK37763 DNA seque
C	20	20	66.7	1328	6 ABA02964 Human rib
21	20	66.7	9058	2	AAV10663 Mouse Fas
22	20	66.7	9058	4	AAH87652 Mouse liv
23	20	66.7	9058	6	AAH49583 Murine tu

24	20	66.7	9058	6	ABQ75766	Murine ca	
25	20	66.7	34769	4	AAH46774	Tumour su	
26	19.6	65.3	47	3	AAZ67933	Human map	
C	27	19.6	65.3	705	9	ADB49481	Primary r
C	28	19.6	65.3	1491	2	AAQ67726	Comamonas
C	29	19.6	65.3	5823	6	ABK28382	DNA trans
C	30	19.6	65.3	110000	7	AAH52246	Continuation (2 of
C	31	19.4	64.7	203	3	AAH17853	Human sec
C	32	19.4	64.7	639	6	ABQ42337	Oligonuc
C	33	19.4	64.7	639	6	ABQ42336	Oligonuc
C	34	19.4	64.7	945	4	AAH52751	S. epider
C	35	19.4	64.7	2481	8	ADA32102	DNA encod
C	36	19.4	64.7	2956	4	AAH54693	S. epider
C	37	19.4	64.7	3543	4	AAH54720	S. epider
C	38	19.4	64.7	3727	7	ADA53054	Human cod
C	39	19.4	64.7	5216	6	ABL32346	Human imm
C	40	19.4	64.7	5216	6	ABL34458	Human met
C	41	19.4	64.7	6314	6	ABL54314	Chemical
C	42	19.4	64.7	7843	6	ABK31504	Signal tr
C	43	19.4	64.7	7843	6	ABL70475	Chemical
C	44	19.4	64.7	7843	6	AAH61438	Human gen
C	45	19.4	64.7	8423	6	ABL33406	Human imm

ALIGNMENTS

RESULT 1
ACD28584
ID ACD28584 standard; DNA; 30 BP.
XX
AC ACD28584;
XX
DT 10-OCT-2003 (first entry)
XX
DE E. coli Arsr binding oligonucleotide PLASLIT.
XX
KW Arsr; arsenic resistance operon; biosensor; ss; arsenic.
XX
OS Escherichia coli.
XX
PW US2003096275-A1.
XX
PD 22-MAY-2003.
XX
PF 15-AUG-2002; 2002US-00222952.
XX
PR 20-AUG-2001; 2001US-0313714P.
XX
PA (LAIN/) LAING L G.
XX
PI Laing LG;
XX
DR WPI; 2003-576876/54.
XX
PT New system comprising isolated protein and nucleic acid, and a detection system that indicates a change in binding of the protein to the nucleic acid in the presence of the analyte, useful for detecting the presence of analyte in a sample.
XX
PS Claim 35; Page 15; 36pp; English.
XX
CC The invention relates to a new system (biosensor) for detecting the presence of analyte in a sample comprising: (1) an isolated protein that specifically binds the analyte; (2) an isolated nucleic acid containing a specific binding sequence that is bound specifically by the protein; and (3) a detection system that indicates a change in binding of the protein to the nucleic acid in the presence of the analyte. Also included are a biosensor device for detecting the presence of an analyte in a sample, detecting the presence of an analyte in a sample, and an Arsr (encoded by part of the arsenic resistance operon of E. coli) protein comprising an amino acid sequence that is at least 90% identical to amino acids 1-97 of the Arsr sequence appearing as ABU63440 binding to a nucleic acid

CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
CC system is useful for detecting the presence of analyte in a sample. The
CC present sequence is the top strand of an oligonucleotide which binds to a
CC plasmid expressed ArsR protein and is used in the biosensor of the
CC invention
XX
SQ Sequence 30 BP; 5 A; 2 C; 7 G; 16 T; 0 U; 0 Other;

Query Match 100.0%; Score 30; DB 8; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTAATCATATCGCTTTTGGTTATGTGTTG 30
Db 1 TTAATCATATCGCTTTTGGTTATGTGTTG 30

RESULT 2

ACD28585/C
ID ACD28585 standard; DNA; 28 BP.

AC ACD28585;

XX 10-OCT-2003 (first entry)

DE E. coli ArsR binding oligonucleotide PLASLIB.

XX ArsR; arsenic resistance operon; biosensor; ss; arsenic.

OS Escherichia coli.

XX US2003096275-A1.

XX 22-MAY-2003.

XX 15-AUG-2002; 2002US-00222952.

XX 20-AUG-2001; 2001US-0313714P.

XX (LAIN/) LAING L G.

XX Laing LG;

XX WPI; 2003-576876/54.

XX New system comprising isolated protein and nucleic acid, and a detection
PT system that indicates a change in binding of the protein to the nucleic
PT acid in the presence of the analyte, useful for detecting the presence of
PT analyte in a sample.

XX Claim 35; Page 15; 36pp; English.

XX The invention relates to a new system (biosensor) for detecting the
CC presence of analyte in a sample comprising: (1) an isolated protein that
CC specifically binds the analyte; (2) an isolated nucleic acid containing a
CC specific binding sequence that is bound specifically by the protein; and
CC (3) a detection system that indicates a change in binding of the protein
CC to the nucleic acid in the presence of the analyte. Also included are a
CC biosensor device for detecting the presence of an analyte in a sample,
CC detecting the presence of an analyte in a sample, and an ArsR (encoded by
CC part of the arsenic resistance operon of E. coli) protein comprising an
CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
CC the ArsR sequence appearing as AB063440 binding to a nucleic acid
CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
CC system is useful for detecting the presence of analyte in a sample. The
CC present sequence is the bottom strand of an oligonucleotide which binds
CC to a plasmid expressed ArsR protein and is used in the biosensor of the
CC invention

SQ Sequence 28 BP; 14 A; 7 C; 2 G; 5 T; 0 U; 0 Other;

Query Match 93.3%; Score 28; DB 8; Length 28;
Best Local Similarity 100.0%; Pred. No. 0.15;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 AATCATATCGCTTTTGGTTATGTGTTG 30
Db 28 AATCATATCGCTTTTGGTTATGTGTTG 1

RESULT 3

ACD28588
ID ACD28588 standard; DNA; 23 BP.

XX ACD28588;

XX 10-OCT-2003 (first entry)

DE E. coli ArsR binding oligonucleotide PLASLIB.

XX ArsR; arsenic resistance operon; biosensor; ss; arsenic.

OS Escherichia coli.

XX US2003096275-A1.

XX 22-MAY-2003.

XX 15-AUG-2002; 2002US-00222952.

XX 20-AUG-2001; 2001US-0313714P.

XX (LAIN/) LAING L G.

XX Laing LG;

XX WPI; 2003-576876/54.

XX New system comprising isolated protein and nucleic acid, and a detection
PT system that indicates a change in binding of the protein to the nucleic
PT acid in the presence of the analyte, useful for detecting the presence of
PT analyte in a sample.

XX Claim 35; Page 15; 36pp; English.

XX The invention relates to a new system (biosensor) for detecting the
CC presence of analyte in a sample comprising: (1) an isolated protein that
CC specifically binds the analyte; (2) an isolated nucleic acid containing a
CC specific binding sequence that is bound specifically by the protein; and
CC (3) a detection system that indicates a change in binding of the protein
CC to the nucleic acid in the presence of the analyte. Also included are a
CC biosensor device for detecting the presence of an analyte in a sample,
CC detecting the presence of an analyte in a sample, and an ArsR (encoded by
CC part of the arsenic resistance operon of E. coli) protein comprising an
CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
CC the ArsR sequence appearing as AB063440 binding to a nucleic acid
CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
CC system is useful for detecting the presence of analyte in a sample. The
CC present sequence is the top strand of an oligonucleotide which binds to a
CC plasmid expressed ArsR protein and is used in the biosensor of the
CC invention

SQ Sequence 23 BP; 5 A; 2 C; 4 G; 12 T; 0 U; 0 Other;

Query Match 76.7%; Score 23; DB 8; Length 23;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTAATCATATCGCTTTTGGTTA 23
Db 1 TTAATCATATCGCTTTTGGTTA 23

RESULT 4

ABK28117
ID ABK28137 standard; DNA; 6391 BP.


```

Best Local Similarity 85.7%; Pred. No. 55;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 TAATCATATGCGTTTTCGTTATGTT 29
DB 18848 TAATCATATGCGTTTTCGTTATGTT 18821

RESULT 6
AAA81490_14/c
Continuation (15 of 15) of AAA81490 from base 1400001 (N. meningitidis B full length gen
WP Sequence split into 15 fragments LOCUS AAA81490 Accession Aaa81490
WP Fragment Name Begin End
WP AAA81490_00 1 110000
WP AAA81490_01 100001 210000
WP AAA81490_02 200001 310000
WP AAA81490_03 300001 410000
WP AAA81490_04 400001 510000
WP AAA81490_05 500001 610000
WP AAA81490_06 600001 710000
WP AAA81490_07 700001 810000
WP AAA81490_08 800001 910000
WP AAA81490_09 900001 1010000
WP AAA81490_10 1000001 1110000
WP AAA81490_11 1100001 1210000
WP AAA81490_12 1200001 1310000
WP AAA81490_13 1300001 1410000
WP AAA81490_14 1400001 1437668

Query Match 72.0%; Score 21.6; DB 3; Length 37668;
Best Local Similarity 85.7%; Pred. No. 67;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 TAATCATATGCGTTTTCGTTATGTT 29
DB 9388 TAATCATATGCGTTTTCGTTATGTT 9361

RESULT 7
AAA81490_13/c
Continuation (14 of 15) of AAA81490 from base 1300001 (N. meningitidis B full length gen
WP Sequence split into 15 fragments LOCUS AAA81490 Accession Aaa81490
WP Fragment Name Begin End
WP AAA81490_00 1 110000
WP AAA81490_01 100001 210000
WP AAA81490_02 200001 310000
WP AAA81490_03 300001 410000
WP AAA81490_04 400001 510000
WP AAA81490_05 500001 610000
WP AAA81490_06 600001 710000
WP AAA81490_07 700001 810000
WP AAA81490_08 800001 910000
WP AAA81490_09 900001 1010000
WP AAA81490_10 1000001 1110000
WP AAA81490_11 1100001 1210000
WP AAA81490_12 1200001 1310000
WP AAA81490_13 1300001 1410000
WP AAA81490_14 1400001 1437668

Query Match 72.0%; Score 21.6; DB 3; Length 110000;
Best Local Similarity 85.7%; Pred. No. 72;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 TAATCATATGCGTTTTCGTTATGTT 29
DB 109398 TAATCATATGCGTTTTCGTTATGTT 109361

RESULT 8
AAF21610/c
ID AAF21610 standard; DNA; 349980 BP.
XX AC AAF21610;

```

```

XX 13-MAR-2001 (first entry)
XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:111.
XX DE Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
XX KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
XX KW ds.
XX OS Neisseria meningitidis.
XX PN WO200066791-A1.
XX PD 09-NOV-2000.
XX PF 08-MAR-2000; 2000WO-US005928.
XX PR 30-APR-1999; 99US-0132068P.
XX PR 08-OCT-1999; 99WO-US023573.
XX PR 28-FEB-2000; 2000GB-00004695.
XX PA (CHIR ) CHIRON CORP.
XX PA (GENO-) INST GENOMIC RES.
XX PI Pizza M, Hickey B, Peterson J, Tettelin H, Venter JC;
XX PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
XX PI Rappuoli R, Frazer CM, Grandi G;
XX DR WPI; 2000-647603/62.
XX Neisseria meningitidis B full length genome sequence and open reading
XX PT frames are used to detect, treat and prevent Neisserial infections.
XX PS Claim 7; Appendix A; 692pp; English.
XX CC The present invention describes the full length genome of Neisseria
XX CC meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613
XX CC represent fragments of the NMB genomic sequence, as the sequence was too
XX CC long to go in a record on its own it was split into 8 sequences which
XX CC overlap each other at the beginning and end of each sequence by 49980 bp
XX CC (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of
XX CC AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of
XX CC AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins
XX CC given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR
XX CC primers which are used in the exemplification of the present invention.
XX CC The NMB genome and fragments from it have antibacterial activity, and can
XX CC be used in vaccines and gene therapy. Neisseria nucleic acids, proteins
XX CC and/or antibodies which binds to the proteins can be used in compositions
XX CC for treating or preventing infection due to Neisserial bacteria or as a
XX CC diagnostic reagent for detecting the presence of Neisserial bacteria or
XX CC of antibodies raised to Neisserial bacteria. Computers, computer memory,
XX CC computer storage medium or computer databases can be used in a search to
XX CC identify open reading frames (ORFs) or coding sequences within the NMB
XX CC genome. The DNA sequences provide further opportunities to find antigenic
XX CC or immunogenic proteins which are more effective in vaccines than the
XX CC outer membrane proteins currently used
XX SQ Sequence 349980 BP; 86771 A; 92803 C; 86340 G; 84066 T; 0 U; 0 Other;

Query Match 72.0%; Score 21.6; DB 3; Length 349980;
Best Local Similarity 85.7%; Pred. No. 77;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 TAATCATATGCGTTTTCGTTATGTT 29
DB 239000 TAATCATATGCGTTTTCGTTATGTT 238973

RESULT 9
ABL87877
ID ABL87877 standard; cDNA; 343 BP.
XX AC ABL87877;

```

XX 17-MAY-2002 (first entry)
 XX Human ovarian cancer related cDNA clone SEQ ID NO:10855.
 XX KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
 XX OS Homo sapiens.

XX WO200192581-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US017755.

XX 26-MAY-2000; 2000US-0207484P.

XX (CORI-) CORIXA CORP.

XX Algate PA, Harlocker SL, Jones R;

XX WPI; 2002-122075/16.

XX Composition for therapy and diagnosis of ovarian cancer comprising
 PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
 PT polypeptide, antibody specific to polypeptide or T cell expressing
 PT polypeptide.

PS Claim 1; SEQ ID NO 10855; 489pp; English.

XX The present invention describes a composition (I) comprising: carriers
 CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
 CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
 CC from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
 CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
 CC or antigen presenting cells that express (II). (I) has cytostatic
 CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
 CC detecting ovarian cancer in a patient's biological sample preferably
 CC serum or ovarian tissue. The method comprises contacting a biological
 CC sample from a patient with (IV), detecting the amount of polynucleotide
 CC hybridising to (IV) and comparing the amount to a predetermined cutoff
 CC value and thereby detecting ovarian cancer in the patient, where the
 CC amount of polynucleotide hybridising to (IV) is detected preferably by
 CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
 CC useful for stimulating and/or expanding T cells specific for an ovarian
 CC tumour protein comprising contacting T cells with (III) or (II). (III) is
 CC useful in design and preparation of ribozyme molecules for inhibiting
 CC expression of the tumour polypeptides and proteins in tumour cells; and
 CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
 CC library using well known techniques

XX Sequence 343 BP; 117 A; 52 C; 50 G; 124 T; 0 U; 0 Other;

Query Match 70.7%; Score 21.2; DB 6; Length 343;
 Best Local Similarity 88.5%; Pred. No. 72;
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 TTATCATATGCGTTTTCGTTATGCT 26

DB 78 TTATCAGATGCGCTTTTAGTTATGCT 103

RESULT 10

ACD28589/c
 ID ACD28589 standard; DNA; 21 BP.

XX ACD28589;

XX 10-OCT-2003 (first entry)

XX E. coli ArsR binding oligonucleotide PLASS1B.

XX ArsR; arsenic resistance operon; biosensor; ss; arsenic.

XX Escherichia coli.
 XX US2003096275-A1.
 XX 22-MAY-2003.

XX 15-AUG-2002; 2002US-00222952.

XX 20-AUG-2001; 2001US-0313714P.

XX (LAIN/) LAING L G.

XX Laing LG;

XX WPI; 2003-576876/54.

XX New system comprising isolated protein and nucleic acid, and a detection
 PT system that indicates a change in binding of the protein to the nucleic
 PT acid in the presence of the analyte, useful for detecting the presence of
 PT analyte in a sample.

XX Claim 35; Page 15; 36pp; English.

XX The invention relates to a new system (biosensor) for detecting the
 CC presence of analyte in a sample comprising: (1) an isolated protein that
 CC specifically binds the analyte; (2) an isolated nucleic acid containing a
 CC specific binding sequence that is bound specifically by the protein; and
 CC (3) a detection system that indicates a change in binding of the protein
 CC to the nucleic acid in the presence of the analyte. Also included are a
 CC biosensor device for detecting the presence of an analyte in a sample,
 CC detecting the presence of an analyte in a sample, and an ArsR (encoded by
 CC part of the arsenic resistance operon of E. coli) protein comprising an
 CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
 CC the ArsR sequence appearing as ABU63440 binding to a nucleic acid
 CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the bottom strand of an oligonucleotide which binds
 CC to a plasmid expressed ArsR protein and is used in the biosensor of the
 CC invention

XX Sequence 21 BP; 10 A; 4 C; 2 G; 5 T; 0 U; 0 Other;

Query Match 70.0%; Score 21; DB 8; Length 21;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 AATCATATGCGTTTTCGTTA 23

DB 21 AATCATATGCGTTTTCGTTA 1

RESULT 11

ADA29013

ID ADA29013 standard; DNA; 459 BP.

XX ADA29013;

XX 20-NOV-2003 (first entry)

XX DNA encoding Acinetobacter baumannii protein #300.

XX ds; Gene; Acinetobacter baumannii; bacterial disease; antibacterial;
 KW vaccine; plant biocontrol agent.

XX Acinetobacter baumannii.

XX US6562958-B1.

XX 13-MAY-2003.

XX 04-JUN-1999; 99US-00328352.


```

PR 09-JUN-1998; 98US-0088701P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Breton G, Bush D;
XX WPI; 2003-576092/54.
XX P-PSDB; ADA31139.
XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX Example; SEQ ID NO 300; 328pp; English.
XX The invention relates to isolated Acinetobacter baumannii nucleic acids.
XX The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents DNA encoding an A. baumannii
CC protein.
XX SQ Sequence 459 BP; 93 A; 90 C; 102 G; 174 T; 0 U; 0 Other;
Query Match 68.7%; Score 20.6; DB 8; Length 459;
Best Local Similarity 85.2%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TTAATCATATCGGTTTGGTTATGTCG 27
Db 72 TTATTCAGTTGCGTTTGGTTATGTCG 98

RESULT 12
ABLI6815
ID ABLI6815 standard; DNA; 1137 BP.
XX AC ABLI6815;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 1918.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS pharmacological; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX XX 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX XX (PEXE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 1918; 21pp + Sequence Listing; English.

```

```

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceuticals. The invention
CC discloses genomic DNA sequences (ABLI6176-ABL30511), expressed DNA
CC sequences (ABLI01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1137 BP; 309 A; 239 C; 262 G; 327 T; 0 U; 0 Other;
Query Match 68.0%; Score 20.4; DB 4; Length 1137;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 TTAATCATATCGGTTTGGTTATGTCGTCG 30
Db 203 TTATACATATCGGTTTGGTTAAGTCGTCG 232

RESULT 13
AAF56298
ID AAF56298 standard; DNA; 1543 BP.
XX AC AAF56298;
XX DT 18-APR-2001 (first entry)
XX DE IS2 element sequence #5.
XX KW Miniature inverted repeat transposable element; MITE; duplication;
XX OS regulate; IS2; ds.
XX OS Daucus carota.
XX PN WO200105986-A2.
XX PD 25-JAN-2001.
XX PF 19-JUL-2000; 2000WO-JF004837.
XX XX 21-JUL-1999; 99JP-00206316.
XX PR 21-JUL-1999; 99JP-00206320.
XX PR 12-JUN-2000; 2000JP-00175825.
XX XX (SANE-) SAN-EI GEN FFI INC.
XX PA (OZEK/) OZEKI Y.
XX PI Ozeki Y, Oyanagi M, Fukuda T, Koda T;
XX WPI; 2001-147351/15.
XX Novel miniature inverted-repeat transposable element (MITE)-like element
PT useful for constructing transgene expression cassette for stably creating
PT genetically engineered organisms capable of expression of transgene.
XX Disclosure; Page 103-104; 104pp; English.
XX The present invention relates to a miniature inverted-repeat transposable
CC element (MITE)-like element capable of causing duplication of a target
CC sequence at the site of its insertion in a genomic gene. The invention is
CC useful for causing expression of a transgene in a plant. It can also be
CC used for inducing or regulating the expression in a plant of the gene
CC introduced
XX SQ Sequence 1543 BP; 569 A; 244 C; 214 G; 516 T; 0 U; 0 Other;
Query Match 68.0%; Score 20.4; DB 4; Length 1543;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```


Search completed: May 26, 2004, 17:49:59
Job time : 219.231 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 16:21:09 ; Search time 2078.59 Seconds
(without alignments)
430.997 Million cell updates/sec

Title: US-10-676-299-3

Perfect score: 30

Sequence: 1 ttaatcatatgcgttttggttatgttg 30

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513239 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hic:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_nam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23.2	77.3	677	12	BG845098
2	22	73.3	400	28	BH370025
3	22	73.3	695	14	CD845514
4	22	73.3	814	28	BH376658

ALIGNMENTS

RESULT 1
BG845098

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

BG845098 677 bp mRNA linear EST 29-MAY-2001
1024008G02.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
Chlamydomonas reinhardtii cDNA, mRNA sequence.

BG845098

BG845098.1 GI:14226282

EST.

Chlamydomonas reinhardtii

Chlamydomonas reinhardtii

Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

Chlamydomonadaceae; Chlamydomonas.

1 (bases 1 to 677)

Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,

McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.

Analyses of the Chlamydomonas reinhardtii Genome: A Model,

Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants; project phase 2

Unpublished (2000)

Contact: Charles Hauser

DCMB Box 91000

Duke University

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu.

Location/Qualifiers

5	21.6	72.0	497	28	AZ248907
6	21.6	72.0	666	14	CD843195
7	21.2	70.7	154	13	BQ704555
8	21.2	70.7	343	9	BQ704555
9	21.2	70.7	359	13	AI738836
10	21	70.7	359	13	AI738836
11	21	70.0	165	28	BH193131
12	21	70.0	220	10	BH193131
13	21	70.0	222	10	BH193131
14	21	70.0	306	14	CA907972
15	21	70.0	309	14	CA907971
16	21	70.0	430	14	CA907970
17	21	70.0	500	12	BH78636
18	21	70.0	539	12	BH78636
19	21	70.0	544	9	AI666913
20	21	70.0	550	10	AI666913
21	21	70.0	582	12	BM857902
22	21	70.0	587	12	BM857902
23	21	70.0	601	12	BM857902
24	21	70.0	602	12	BM857902
25	21	70.0	608	12	BM857902
26	21	70.0	611	28	AG785876
27	21	70.0	701	29	AG173299
28	21	70.0	723	14	CD842667
29	21	70.0	863	28	BH135394
30	20.6	68.7	309	12	AK043740
31	20.6	68.7	309	12	AK043740
32	20.6	68.7	670	28	BH989212
33	20.6	68.7	768	12	BM171079
34	20.6	68.7	862	29	C9119993
35	20.4	68.0	919	29	CNS017RV
36	20.4	68.0	187	9	AV228225
37	20.4	68.0	219	12	BG586110
38	20.4	68.0	357	28	BZ094929
39	20.4	68.0	422	14	CD845285
40	20.4	68.0	439	28	AZ283390
41	20.4	68.0	484	10	AW585883
42	20.4	68.0	567	28	AZ699341
43	20.4	68.0	575	9	AI260326
44	20.4	68.0	586	29	CC773739
45	20.4	68.0	603	28	BH187898
	20.4	68.0	603	29	CNS0751X

AZ248907 RPCI-23-5
CD843195 RPO2.130P
BQ704555 Bn01.02b0
AI738836 cr28g11.x
BH193131 BX11.1464
BH193131 BX11.1464
BF813601 MR2-CI018
BH186053 BH186053
CA907972 PCS20237
CA907971 PCS18727
CA907970 PCS03468
BH78636 fl80h07.x
BH186053 fl80h07.x
AI666913 fd16e03.x
AW344027 fl72f10.x
BM857902 fy37d07.x
BG307569 fl59h08.x
BM184130 fv88d01.x
BM095920 fv28d08.x
BM183874 fv64g12.x
AG785876 HS_3053_A
AG173299 Pan trogl
CD842667 RPO2.1281
BH135394 RNTNC30TR
AK043740 Mus muscu
BM164586 EST567109
BH989212 Oe991C01.
BM171079 EST573602
C9119993 MBED013TF
AL108421 Drosophil
AV228225 AV228225
BG586110 EST487875
BZ094929 CH230-214
CD845285 RPO2.1421
AZ283390 RPCI-23-1
AW585883 EST317506
AZ699341 RPCI-23-2
AI260326 LP04121.5
CC773739 CH240.33A
BH187898 O3*_J22-
AL624842 T3 end of

```

1. 677
source
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
II."
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP (acetate-containing) medium in the
light, TAP medium in the dark, HS (minimal) medium in
ambient levels of CO2 and HS medium bubbled with 5% CO2.
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with EXassist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."

ORIGIN
Query Match 77.3%; Score 23.2; DB 12; Length 677;
Best Local Similarity 89.3%; Pred. No. 4.8e+02;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AATCATATGCGTTTGGTTATGTGTTG 30
Db 460 AGTCATATGCGTCTTTGTTATGTGTTG 487

RESULT 2
BH370025
LOCUS
DEFINITION
AG-ND-170H24.TR ND-TAM Anopheles gambiae genomic clone
AG-ND-170H24, genomic survey sequence.
ACCESSION
BH370025
VERSION
BH370025.1 GI:17316128
KEYWORDS
GSS.
SOURCE
Anopheles gambiae (African malaria mosquito)
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE
1 (bases 1 to 400)
Kong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J.,
Ren,G., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B.,
Gardner,M.J. and Collins,F.H.
Construction of a BAC library and generation of BAC end
sequence-tagged connectors for genome sequencing of the African
malaria mosquito Anopheles gambiae
Mol. Genet. Genomics 268 (6), 720-728 (2003)
25542063
12655398
COMMENT
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2125, USA using a HindIII
partial digest.
Seq primer: M13 Rev
Class: BAC ends.

```

ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.

REFERENCE 1 (bases 1 to 814)
AUTHORS Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren, C., Huff, E.R., Carlisle, J.L., Black, K., Zhang, H.-B., Gardner, M.J., and Collins, F.H.

TITLE Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito Anopheles gambiae

JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)
MEDLINE 22542063
PUBMED 12655398

COMMENT Other GSSs: AG-ND-171113.TP
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@igr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 Rev
Class: BAC ends.

FEATURES
source
1..814
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-171113"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

ORIGIN
Query Match 73.3%; Score 22; DB 28; Length 814;
Best Local Similarity 83.3%; Pred. No. 1.1e+03;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTGGTTATGTTG 30
Db 306 TAAATCAATGCGTTTGGTTATGTTATG 335

RESULT 5
AZ248907
LOCUS AZ248907
DEFINITION RPCI-23-57D7 TJ RPCI-23 Mus musculus genomic clone RPCI-23-57D7, genomic survey sequence.
ACCESSION AZ248907
VERSION AZ248907.1 GI:8562110
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 497)
AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akirret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-57D7.TV
Contact: Shaying Zhao

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: schae@igr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@jlong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (inforesgen.com). BAC end page: http://www.tigr.org/tbdr/bac_ends/mouse/bac_end_intro.html
plate: 57 row: D column: 7
Seq primer: SP6
Class: BAC ends.

FEATURES
source
1..497
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clones="RPCI-23-57D7"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN
Query Match 72.0%; Score 21.6; DB 28; Length 497;
Best Local Similarity 85.7%; Pred. No. 1.6e+03;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TTAATCATATGCGTTTGGTTATGTTG 28
Db 111 TCAATCATATGCGTTTGGTTATGTTG 138

RESULT 6
CD843195
LOCUS CD843195
DEFINITION RFO2.130P04P011011 RFO2 Brassica napus cDNA clone RFO2130P04, mRNA sequence.
ACCESSION CD843195
VERSION CD843195.1 GI:32525135
KEYWORDS EST.
SOURCE Brassica napus (rape)
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 666)
AUTHORS Genopiante.
TITLE Genopiante, a major partnership french program in plant genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Genopiante
Genopiante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french plant genomics programme 'Genopiante' (<http://www.genopiante.com>) and <http://genopiante-info.infobiogen.fr>.
Location/Qualifiers
1..666
/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="samourai (restored line)"

LOCUS	AI738836	343 bp	mRNA	linear	EST 18-JUN-1999
DEFINITION	tr28911.x1 NCI_CGAP_Ov23 Homo sapiens cDNA clone IMAGE:2219684 3'				
ACCESSION	AI738836				
VERSION	AI738836.1				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 343)				
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaps-r@mail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Sequencing by: Washington University Genome Sequencing Center DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html Seq primer: -400P from Gibco High quality sequence stop: 337.				
FEATURES	<p>Location/Qualifiers</p> <p>1..343</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="IMAGE:2219684"</p> <p>/tissue_type="tumor, 5 pooled (see description)"</p> <p>/lab_host="DH10B"</p> <p>/clone_lib="NCI CGAP Ov23"</p> <p>/note="Organ: ovary; Vector: pCMV-Sport6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.35 kb. Tumor types include: mixed Mullerian tumor, papillary serous, clear cell, spindle cell. All are primary tumors, metastasis positive. Life Technologies catalog #: 11534-013"</p>				
ORIGIN	<p>Query Match 70.7%; Score 21.2; DB 9; Length 343;</p> <p>Best Local Similarity 88.5%; Pred. No. 2.4e+03;</p> <p>Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;</p>				
Qy	1 TTAATCATATGCGTTTGGTTATGT 26				
Db	78 TTAATCATATGCGTTTGGTTATGT 103				
RESULT 9	<p>EX111464</p>				
LOCUS	EX111464 NCI_CGAP Ov23 Homo sapiens cDNA clone IMAGE:2219684, mRNA sequence.				
DEFINITION	IMAGES:2219684, mRNA sequence.				
ACCESSION	BX111464				
VERSION	BX111464.1				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 359)				
TITLE	Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M., Radelof, U., Schneider, D. and Korn, B.				
JOURNAL	Human Unigeneset - RZPD				
COMMENT	Unpublished (2003) Contact: Ina Rolfs RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany				

LOCUS	BQ704555	154 bp	mRNA	linear	EST 16-JUL-2002
DEFINITION	Bn01_02b08 A				
ACCESSION	Bn01_02b08 A				
VERSION	Bn01_02b08 A				
KEYWORDS	EST.				
SOURCE	Brassica napus				
ORGANISM	Brassica napus (rape)				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.				
AUTHORS	1 (bases 1 to 154)				
TITLE	Singh, J., Allard, G., Tinker, N., Robert, L., Lacroix, C., De Moors, A., Chagnon, J., Parah, S., Couroux, P. and Hattori, J.				
JOURNAL	Expressed Sequence Tags from constitutively frost tolerant transgenic Brassica napus overexpressing BNCBF17				
COMMENT	Unpublished (2002) Contact: Singh, J.A. Eastern Cereal and Oilseed Research Centre Agriculture and Agri-food Canada KW Neatby Bldg., Central Experimental Farm, Ottawa, Ontario, K1A 0C6, Canada Tel: (613) 759-1662 Fax: (613) 759-1701 Email: singhja@agr.ca.				
FEATURES	<p>Location/Qualifiers</p> <p>1..154</p> <p>/organism="Brassica napus"</p> <p>/mol_type="mRNA"</p> <p>/cultivar="Westar"</p> <p>/db_xref="taxon:3708"</p> <p>/clone="Bn01_02b08"</p> <p>/tissue_type="fourth leaf"</p> <p>/dev_stage="3 weeks seedling grown at room temperature"</p> <p>/clone_lib="Bn01_AAPC_ECORC transgenic Brassica napus over expressing BNCBF17 constitutively frost tolerant"</p> <p>/note="Vector: Bluescript SK+/XhoI-EcoRI; Site 1: EcoRI; Site 2: XhoI; Germinated in soil flats and seedlings grown for 3 weeks in a Conviron E-15 cabinet set at 20oC/16 hr light (250 Em-2sec-1) and 16 oC / 8 hr dark. Fourth leaves collected at 9 am and immediately frozen."</p>				
ORIGIN	<p>Query Match 70.7%; Score 21.2; DB 13; Length 154;</p> <p>Best Local Similarity 82.1%; Pred. No. 2.7e+03;</p> <p>Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;</p>				
Qy	3 AATCATATGCGTTTGGTTATGTG 30				
Db	94 RATCATATGCGTTTGGTTATGTG 67				
RESULT 8	<p>AI738836</p>				

LOCUS	BQ704555/c	154 bp	mRNA	linear	EST 16-JUL-2002
DEFINITION	Bn01_02b08 A				
ACCESSION	Bn01_02b08 A				
VERSION	Bn01_02b08 A				
KEYWORDS	EST.				
SOURCE	Brassica napus				
ORGANISM	Brassica napus (rape)				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.				
AUTHORS	1 (bases 1 to 154)				
TITLE	Singh, J., Allard, G., Tinker, N., Robert, L., Lacroix, C., De Moors, A., Chagnon, J., Parah, S., Couroux, P. and Hattori, J.				
JOURNAL	Expressed Sequence Tags from constitutively frost tolerant transgenic Brassica napus overexpressing BNCBF17				
COMMENT	Unpublished (2002) Contact: Singh, J.A. Eastern Cereal and Oilseed Research Centre Agriculture and Agri-food Canada KW Neatby Bldg., Central Experimental Farm, Ottawa, Ontario, K1A 0C6, Canada Tel: (613) 759-1662 Fax: (613) 759-1701 Email: singhja@agr.ca.				
FEATURES	<p>Location/Qualifiers</p> <p>1..154</p> <p>/organism="Brassica napus"</p> <p>/mol_type="mRNA"</p> <p>/cultivar="Westar"</p> <p>/db_xref="taxon:3708"</p> <p>/clone="Bn01_02b08"</p> <p>/tissue_type="fourth leaf"</p> <p>/dev_stage="3 weeks seedling grown at room temperature"</p> <p>/clone_lib="Bn01_AAPC_ECORC transgenic Brassica napus over expressing BNCBF17 constitutively frost tolerant"</p> <p>/note="Vector: Bluescript SK+/XhoI-EcoRI; Site 1: EcoRI; Site 2: XhoI; Germinated in soil flats and seedlings grown for 3 weeks in a Conviron E-15 cabinet set at 20oC/16 hr light (250 Em-2sec-1) and 16 oC / 8 hr dark. Fourth leaves collected at 9 am and immediately frozen."</p>				
ORIGIN	<p>Query Match 70.7%; Score 21.2; DB 13; Length 154;</p> <p>Best Local Similarity 82.1%; Pred. No. 2.7e+03;</p> <p>Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;</p>				
Qy	3 AATCATATGCGTTTGGTTATGTG 30				
Db	94 RATCATATGCGTTTGGTTATGTG 67				
RESULT 8	<p>AI738836</p>				

```

RZPD; IMAGE998L215489.
RZPDLIB; I.M.A.G.E. CDNA Clone Collection;
Human Unigeneset - RZPD3 (RZPDLIB No.972)
http://www.rzpd.de/CloneCards/cgi-
bin/showlib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Keubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
ML3u, Primer sequence: CCGTGTAAACGACGCCAGT.
FEATURES
    source
        1..359
            Location/Qualifiers
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE998L215489 ; IMAGE:2219684"
                /tissue_type="tumor, 5 pooled (see description)"
                /lab_host="BH10B"
                /clone_lib="NCI_CGAP_Ov23"
                /note="Organ: ovary; Vector: pCMV-SPORT6; Site 1: SalI;
                Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr.
                Average insert size 1.35 kb. Tumor types include: mixed
                Mullerian tumor, papillary serous, clear cell, spindle
                cell. All are primary tumors, metastasis positive. Life
                Technologies catalog #: 11534-013"
ORIGIN
    Query Match      70.7%; Score 21.2; DB 13; Length 359;
    Best Local Similarity 88.5%; Pred. No. 2.3e+03;
    Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

    QY 1 TTAATCATATGCGTTTTCGTTATGT 26
    Db 79 TTAATCAGAGCGCTTTTAGTTATGT 104

RESULT 10
BH193131
LOCUS BH193131
DEFINITION TC3-73D19.TP TC3 Trypanosoma cruzi genomic clone TC3-73D19, genomic
survey sequence.
ACCESSION BH193131
VERSION BH193131.1 GI:16360848
KEYWORDS GSS.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
REFERENCE 1 (bases 1 to 165)
AUTHORS Kluge,S., Edwards,K.E., Nilsson,D., Bontempi,E.J., Myler,P.,
Stuart,K., Ghedin,E., El-Sayed,N.M. and Andersson,B.
Clustering and analysis of BAC-end and GSS sequences from
Trypanosoma cruzi
Unpublished (2001)
Other_GSSs: TC3-73D19.TV
Contact: Bjorn Andersson
Department of Genetics and Pathology
Uppsala University
Rudbeck Laboratory, SE-751 85, Uppsala, Sweden
Tel: 46 18 471 4107
Fax: 46 18 471 4808
Email: bjorn.andersson@genpat.uu.se
Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
TC3. For clone availability, please contact Dr. Bjorn Andersson at
Uppsala University (bjorn.andersson@genpat.uu.se).
Seq primer: SP6
Class: BAC ends.
FEATURES
    source
        1..165
            Location/Qualifiers
                /organism="Trypanosoma cruzi"
                /mol_type="genomic DNA"
                /strain="CL Brener"
                /db_xref="taxon:5693"
                /clone="TC3-73D19"
                /clone_lib="TC3"
                /note="Vector: pBelOBAC11; Site 1: Hin dIII; Constructed
                for Uppsala University by Marie-Christine Le Paslier in
                the laboratory of Denis Le Paslier at the Centre d'Etude
                du Polymorphisme Humain (CEPH), Paris, France. Briefly,
                Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained
                from Dr. Franco da Silveira) was partially digested with
                Hin dIII. High molecular weight fragments were ligated in
                pBelOBAC11 digested with Hin dIII. The average insert
                size is 100 kb. Total clone coverage: approx. 33 X the
                haploid genome."
ORIGIN
    Query Match      70.0%; Score 21; DB 28; Length 165;
    Best Local Similarity 82.8%; Pred. No. 3.1e+03;
    Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

    QY 2 TAAATCATATGCGTTTTCGTTATGTGTTG 30
    Db 3 TCATCGTCTCGCGTTTCGTTATGTGTTG 31

RESULT 11
BF813601
LOCUS BF813601
DEFINITION MR2-C10186-301100-007-b09 C10186 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF813601
VERSION BF813601.1 GI:12144736
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 220)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.P., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jorgensen,C.V.,
O'Hare,M.U., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
{http://www.ludwig.org.br/scripts/gethtml2.pl?l=MR2&t2=MR2-C10186-
301100-007-b09&t3=2000-11-30&t4=1}
Seq primer: puc 18 forward
High quality sequence start: 93
High quality sequence stop: 144.
FEATURES
    source
        1..220
            Location/Qualifiers
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /dev_stage="Adult"
                /clone_lib="C10186"
                /note="Organ: colon_ins; Vector: puc18; Site 1: SmaI;
                Site 2: SmaI; A mini-library was made by cloning products

```


/note="Organ: Suspensor Region of Globular-Stage Embryos; Vector: Triplex2; Site 1: SfilA; Site 2: SfilB; Suspensor regions were micro-dissected from globular-stage embryos six days after pollination from greenhouse-grown plants [Weterings et al., Plant Cell 13, 2409-2425 (2001)]. Double-stranded cDNA was synthesized from suspensor mRNA using the SMART cDNA Library Construction Kit according to the manufacturer (Clontech). The suspensor cDNA fragments were directionally ligated into the SfilI restriction site of the lambda Triplex2 vector (Clontech), and the recombinant cDNAs were transformed into E. coli XL1-Blue cells (Clontech). Suspensor cDNA plasmids used for directional sequencing were obtained by in vivo excision from the lambda Triplex2 recombinants in E. coli BM25.8 cells (Clontech)."

ORIGIN

Query Match 70.0%; Score 21; DB 14; Length 306;
Best Local Similarity 82.8%; Pred. No. 2.8e+03;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 TTAATCATATCGCTTTTGGTTATGTTT 29
|||||
Db 268 TTAATTATATGCTTAATGTTATGTTT 296
|||||

RESULT 14

CA907971

LOCUS

DEFINITION PCS018727 Scarlet Runner Bean Suspensor Region Triplex2 Phaseolus coccineus cDNA 5' similar to Actin, mRNA sequence.

CA907971

VERSION

KEYWORDS

SOURCE

ORGANISM

Phaseolus coccineus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.

REFERENCE

AUTHORS

Rui, A.O., Le, B.H., Weterings, K., Bi, Y.-P., Choi, J.-S., McElroy, K.E., Choi, P.S., Harada, J.J., Fischer, R.L. and Goldberg, R.B.

TITLE Gene Activity in Different Regions of a Post-Fertilization Plant Embryo by EST Analysis

JOURNAL

COMMENT

Unpublished (2002)
Contact: Goldberg, R.B.
Department of Molecular, Cell, & Developmental Biology
University of California, Los Angeles
621 Charles E. Young Drive South, Los Angeles, CA 90095-1606, USA
Tel: 310 825 3270
Fax: 310 825 8201
Email: bobg@ucla.edu
Seq primer: 5' Triplex

FEATURES

source

Location/Qualifiers

1..309
/organism="Phaseolus coccineus"
/mol_type="mRNA"
/cultivar="Hammond's Dwarf Scarlet"
/db_xref="taxon:3886"
/dev_stage="6-days post-pollination"
/clone_lib="Scarlet Runner Bean Suspensor Region Triplex2"
/note="Organ: Suspensor Region of Globular-Stage Embryos; Vector: Triplex2; Site 1: SfilA; Site 2: SfilB; Suspensor regions were micro-dissected from globular-stage embryos six days after pollination from greenhouse-grown plants [Weterings et al., Plant Cell 13, 2409-2425 (2001)]. Double-stranded cDNA was synthesized from suspensor mRNA using the SMART cDNA Library Construction Kit according to the manufacturer (Clontech). The suspensor cDNA fragments were directionally ligated into the SfilI restriction site

of the lambda Triplex2 vector (Clontech), and the recombinant cDNAs were transformed into E. coli XL1-Blue cells (Clontech). Suspensor cDNA plasmids used for directional sequencing were obtained by in vivo excision from the lambda Triplex2 recombinants in E. coli BM25.8 cells (Clontech)."

ORIGIN

Query Match 70.0%; Score 21; DB 14; Length 309;
Best Local Similarity 82.8%; Pred. No. 2.8e+03;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 TTAATCATATCGCTTTTGGTTATGTTT 29
|||||
Db 270 TTAATTATATGCTTAATGTTATGTTT 298
|||||

RESULT 15

CA907970

LOCUS

DEFINITION PCS03468 Scarlet Runner Bean Suspensor Region Triplex2 Phaseolus coccineus cDNA 5' similar to Actin, mRNA sequence.

CA907970

VERSION

KEYWORDS

SOURCE

ORGANISM

Phaseolus coccineus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.

1 (bases 1 to 430)

Rui, A.O., Le, B.H., Weterings, K., Bi, Y.-P., Choi, J.-S., McElroy, K.E., Choi, P.S., Harada, J.J., Fischer, R.L. and Goldberg, R.B.

TITLE Gene Activity in Different Regions of a Post-Fertilization Plant Embryo by EST Analysis

JOURNAL

COMMENT

Unpublished (2002)
Contact: Goldberg, R.B.
Department of Molecular, Cell, & Developmental Biology
University of California, Los Angeles
621 Charles E. Young Drive South, Los Angeles, CA 90095-1606, USA
Tel: 310 825 3270
Fax: 310 825 8201
Email: bobg@ucla.edu
Seq primer: 5' Triplex

FEATURES

source

Location/Qualifiers

1..430
/organism="Phaseolus coccineus"
/mol_type="mRNA"
/cultivar="Hammond's Dwarf Scarlet"
/db_xref="taxon:3886"
/dev_stage="6-days post-pollination"
/clone_lib="Scarlet Runner Bean Suspensor Region Triplex2"
/note="Organ: Suspensor Region of Globular-Stage Embryos; Vector: Triplex2; Site 1: SfilA; Site 2: SfilB; Suspensor regions were micro-dissected from globular-stage embryos six days after pollination from greenhouse-grown plants [Weterings et al., Plant Cell 13, 2409-2425 (2001)]. Double-stranded cDNA was synthesized from suspensor mRNA using the SMART cDNA library Construction Kit according to the manufacturer (Clontech). The suspensor cDNA fragments were directionally ligated into the SfilI restriction site of the lambda Triplex2 vector (Clontech), and the recombinant cDNAs were transformed into E. coli XL1-Blue cells (Clontech). Suspensor cDNA plasmids used for directional sequencing were obtained by in vivo excision from the lambda Triplex2 recombinants in E. coli BM25.8 cells (Clontech)."

ORIGIN

Query Match 70.0%; Score 21; DB 14; Length 430;

Best Local Similarity 82.84; Pred. No. 2.6e+03;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 TTAATCATATCGCTTTTGGTTATGCTT 29
Db 338 TTAATTATATGCTTAATTTGGTTATGCTT 366

Search completed: May 26, 2004, 22:26:53
Job time : 2084.59 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OX nucleic - nucleic search, using sw model

Run on: May 26, 2004, 15:22:28 ; Search time 199.949 seconds
(without alignments)
594,900 Million cell updates/sec

Title: US-10-676-299-4
Perfect score: 28
Sequence: 1 caacacataacacaaacgcatacgatt 28

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext i.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseqn 29Jan04:.*
1: Geneseqn1980s:.*
2: Geneseqn1990s:.*
3: Geneseqn2000s:.*
4: Geneseqn2001as:.*
5: Geneseqn2001bs:.*
6: Geneseqn2002s:.*
7: Geneseqn2003as:.*
8: Geneseqn2003bs:.*
9: Geneseqn2003cs:.*
10: Geneseqn2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	28	8	ACD28585 E. coli A
2	28	100.0	30	8	ACD28584 E. coli A
3	21	75.0	21	8	ACD28589 E. coli A
4	21	75.0	23	8	ACD28588 E. coli A
5	20.6	73.6	6391	6	ABK28137 DNA trans
6	20.6	73.6	23532	3	AA881455 N. mening
7	20.6	73.6	37668	3	AA881490_14
8	20.6	73.6	110000	3	AA881490_13
9	20.6	73.6	349980	3	AA881490_11
10	20.6	71.4	1543	3	AA881490_10
11	19.6	70.0	47	3	AA881490_9
12	19.6	70.0	1491	2	AA881490_8
13	19.2	68.6	343	6	ABK28137 DNA trans
14	19.2	68.6	8087	6	ABK28137 DNA trans
15	19.2	68.6	8087	6	ABK28137 DNA trans
16	19.2	68.6	8087	6	ABK28137 DNA trans
17	19.2	68.6	8087	6	ABK28137 DNA trans
18	19.2	68.6	8087	6	ABK28137 DNA trans
19	19.2	68.6	8087	6	ABK28137 DNA trans
20	19.2	68.6	8087	6	ABK28137 DNA trans
21	19.2	68.6	8087	6	ABK28137 DNA trans
22	19.2	68.6	8087	6	ABK28137 DNA trans
23	19.2	68.6	8087	6	ABK28137 DNA trans

AAL49583 Murine tu
 ABQ75766 Murine ca
 AAE33419 DNA encod
 AAL05045 Human rep
 AB197938 Human tes
 ABK40025 Human che
 AAE46774 Tumour su
 Continuation (3 of
 Continuation (13 of
 ADA29013 DNA encod
 ABQ69002 Listeria
 ABQ14038 Oligonuc1
 ABQ14039 Oligonuc1
 ABQ13769 Oligonuc1
 ABQ13768 Oligonuc1
 AC68832 Phototab
 ABQ70524 Listeria
 AB116815 Drosophil
 AAC69498 Human sec
 AB15671 Arabidops
 AB68950 C. neofor
 AB108236 Drosophil

ALIGNMENTS

RESULT 1
 ACD28585
 ID ACD28585 standard; DNA; 28 BP.
 XX AC ACD28585;
 XX AC ACD28585;
 DT 10-OCT-2003 (first entry)
 DE E. coli Arsr binding oligonucleotide PLASL1B.
 KW Arsr; arsenic resistance operon; biosensor; ss; arsenic.
 OS Escherichia coli.
 XX US2003096275-A1.
 XX PD 22-MAY-2003.
 XX PF 15-AUG-2002; 2002US-00222952.
 XX PR 20-AUG-2001; 2001US-0313714P.
 XX (LAIN/) LAING L G.
 PA Laing LG;
 PI WPI; 2003-576876/54.
 DR New system comprising isolated protein and nucleic acid, and a detection
 PT system that indicates a change in binding of the protein to the nucleic
 PT acid in the presence of the analyte, useful for detecting the presence of
 PT analyte in a sample.
 XX Claim 35; Page 15; 36pp; English.
 XX The invention relates to a new system (biosensor) for detecting the
 CC presence of analyte in a sample comprising: (1) an isolated protein that
 CC specifically binds the analyte; (2) an isolated nucleic acid containing a
 CC specific binding sequence that is bound specifically by the protein; and
 CC (3) a detection system that indicates a change in binding of the protein
 CC to the nucleic acid in the presence of the analyte. Also included are a
 CC biosensor device for detecting the presence of an analyte in a sample,
 CC detecting the presence of an analyte in a sample, and an Arsr (encoded by
 CC part of the arsenic resistance operon of E. coli) protein comprising an
 CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
 CC the Arsr sequence appearing as AB063440 binding to a nucleic acid

CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the bottom strand of an oligonucleotide which binds
 CC to a plasmid expressed ArsR protein and is used in the biosensor of the
 CC invention

XX Sequence 28 BP; 14 A; 7 C; 2 G; 5 T; 0 U; 0 Other;

Query Match 100.0%; Score 28; DB 8; Length 28;
 Best Local Similarity 100.0%; Pred. No. 0.07;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAACACATACCAAAACGCATATGATT 28
 |||||
 Db 1 CAACACATACCAAAACGCATATGATT 28

RESULT 2

ACD28584/c
 ID ACD28584 standard; DNA; 30 BP.

XX ACD28584;

XX 10-OCT-2003 (first entry)

DE E. coli ArsR binding oligonucleotide PLASLIT.

KW ArsR; arsenic resistance operon; biosensor; ss; arsenic.

XX Escherichia coli.

XX US2003096275-A1.

XX 22-MAY-2003.

XX 15-AUG-2002; 2002US-00222952.

XX 20-AUG-2001; 2001US-0313714P.

XX (LAIN/) LAING L G.

XX Laing LG;

XX WPI; 2003-576876/54.

XX New system comprising isolated protein and nucleic acid, and a detection
 PT system that indicates a change in binding of the protein to the nucleic
 PT acid in the presence of the analyte, useful for detecting the presence of
 PT analyte in a sample.

XX Claim 35; Page 15; 36pp; English.

XX The invention relates to a new system (biosensor) for detecting the
 CC presence of analyte in a sample comprising: (1) an isolated protein that
 CC specifically binds the analyte; (2) an isolated nucleic acid containing a
 CC specific binding sequence that is bound specifically by the protein; and
 CC (3) a detection system that indicates a change in binding of the protein
 CC to the nucleic acid in the presence of the analyte. Also included are a
 CC biosensor device for detecting the presence of an analyte in a sample,
 CC detecting the presence of an analyte in a sample, and an ArsR (encoded by
 CC part of the arsenic resistance operon of E. coli) protein comprising an
 CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
 CC the ArsR sequence appearing as ABU63440 binding to a nucleic acid
 CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the top strand of an oligonucleotide which binds to a
 CC plasmid expressed ArsR protein and is used in the biosensor of the
 CC invention

XX Sequence 30 BP; 5 A; 2 C; 7 G; 16 T; 0 U; 0 Other;

Query Match 100.0%; Score 28; DB 8; Length 30;
 Best Local Similarity 100.0%; Pred. No. 0.07;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 CAACACATACCAAAACGCATATGATT 28
 |||||
 Db 30 CAACACATACCAAAACGCATATGATT 3

RESULT 3

ACD28589

ID ACD28589 standard; DNA; 21 BP.

XX ACD28589;

XX 10-OCT-2003 (first entry)

XX E. coli ArsR binding oligonucleotide PLASLB.

XX ArsR; arsenic resistance operon; biosensor; ss; arsenic.

XX Escherichia coli.

XX US2003096275-A1.

XX 22-MAY-2003.

XX 15-AUG-2002; 2002US-00222952.

XX 20-AUG-2001; 2001US-0313714P.

XX (LAIN/) LAING L G.

XX Laing LG;

XX WPI; 2003-576876/54.

XX New system comprising isolated protein and nucleic acid, and a detection
 PT system that indicates a change in binding of the protein to the nucleic
 PT acid in the presence of the analyte, useful for detecting the presence of
 PT analyte in a sample.

XX Claim 35; Page 15; 36pp; English.

XX The invention relates to a new system (biosensor) for detecting the
 CC presence of analyte in a sample comprising: (1) an isolated protein that
 CC specifically binds the analyte; (2) an isolated nucleic acid containing a
 CC specific binding sequence that is bound specifically by the protein; and
 CC (3) a detection system that indicates a change in binding of the protein
 CC to the nucleic acid in the presence of the analyte. Also included are a
 CC biosensor device for detecting the presence of an analyte in a sample,
 CC detecting the presence of an analyte in a sample, and an ArsR (encoded by
 CC part of the arsenic resistance operon of E. coli) protein comprising an
 CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
 CC the ArsR sequence appearing as ABU63440 binding to a nucleic acid
 CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the bottom strand of an oligonucleotide which binds
 CC to a plasmid expressed ArsR protein and is used in the biosensor of the
 CC invention

XX Sequence 21 BP; 10 A; 4 C; 2 G; 5 T; 0 U; 0 Other;

Query Match 75.0%; Score 21; DB 8; Length 21;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TAACCAAAACGCATATGATT 28
 |||||
 Db 1 TAACCAAAACGCATATGATT 21

RESULT 4

ACD28588/c

ID ACD28588 standard; DNA; 23 BP.

XX ACD28588;
AC 10-OCT-2003 (first entry)
DT
XX E. coli ArsR binding oligonucleotide PLASS1T.
DE
XX ArsR; arsenic resistance operon; biosensor; ss; arsenic.
KW
XX Escherichia coli.
OS
XX US2003096275-A1.
PN
XX 22-MAY-2003.
PD
XX 15-AUG-2002; 2002US-00222952.
PP
XX 20-AUG-2001; 2001US-0313714P.
PR
XX (LAIN/) LAING L G.
PA
XX Laing LG;
PI
XX WPI; 2003-576876/54.
DR
XX New system comprising isolated protein and nucleic acid, and a detection
PT system that indicates a change in binding of the protein to the nucleic
PT acid in the presence of the analyte, useful for detecting the presence of
PT analyte in a sample.
PT
XX Claim 35, Page 15; 35pp; English.
PS
XX The invention relates to a new system (biosensor) for detecting the
CC presence of analyte in a sample comprising: (1) an isolated protein that
CC specifically binds the analyte; (2) an isolated nucleic acid containing a
CC specific binding sequence that is bound specifically by the protein; and
CC (3) a detection system that indicates a change in binding of the protein
CC to the nucleic acid in the presence of the analyte. Also included are a
CC biosensor device for detecting the presence of an analyte in a sample,
CC detecting the presence of an analyte in a sample, and an ArsR (encoded by
CC part of the arsenic resistance operon of E. coli) protein comprising an
CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
CC the ArsR sequence appearing as ABU63440 binding to a nucleic acid
CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
CC system is useful for detecting the presence of analyte in a sample. The
CC present sequence is the top strand of an oligonucleotide which binds to a
CC plasmid expressed ArsR protein and is used in the biosensor of the
CC invention
XX
SQ Sequence 23 BP; 5 A; 2 C; 4 G; 12 T; 0 U; 0 Other;
Query Match 75.0%; Score 21; DB 8; Length 23;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 TAACCAAAAACGCATATGATT 28
DB 23 TAACCAAAAACGCATATGATT 3
RESULT 5
ID ABK28137/c
XX ABK28137 standard; DNA; 6391 BP.
XX
AC ABK28137;
XX
DT 23-APR-2002 (first entry)
XX
XX DNA transcription associated genomic DNA #6.
DE
XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;

KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW immunological disorder; Werner syndrome; developmental disorder;
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW angiodysplasia; congenital heart disease; HDR syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.
XX
OS Unidentified.
XX WO200192565-A2.
PN
XX 06-DEC-2001.
PD
XX 06-APR-2001; 2001WO-EP003973.
PP
XX 06-APR-2000; 2000DE-01019058.
PR
XX 07-APR-2000; 2000DE-01019173.
PR
XX 30-JUN-2000; 2000DE-01032529.
PR
XX 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX WPI; 2002-090046/12.
DR
XX New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological disorders,
PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or
PT cancer.
XX
XX Claim 1; SEQ ID NO 11; 32pp; English.
PS
XX The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiodysplasia, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification but
CC was obtained in electronic format directly from the European Patent
CC Office
XX
SQ Sequence 6391 BP; 1421 A; 156 C; 1688 G; 3126 T; 0 U; 0 Other;
Query Match 73.6%; Score 20.6; DB 6; Length 6391;
Best Local Similarity 85.2%; Pred. No. 89;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 AACACATAACCAAAAACGCATATGATT 28
DB 1027 AACACATAACCAAAAACGCATATGATT 1001
RESULT 6
ID AAA81455
XX AAA81455 standard; DNA; 23532 BP.
AC AAA81455;

XX 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm_3 SEQ ID NO:3.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;

XX antigen; vaccine; diagnosis; infection; antibacterial; identification;

XX Meningococcus B; MenB; ds.

XX Neisseria meningitidis.

XX WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US023573.

XX 09-OCT-1998; 99US-0103794P.

XX 30-APR-1999; 99US-0132068P.

XX (CHIR) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC; Scarlato V;

XX Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;

XX Rappuoli R, Pizza M;

XX WPI; 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be used

XX in the diagnosis and treatment of N. meningitidis infection and other

XX Neisserial infections, for example, N.gonorrhoea.

XX Claim 7; Page 253-260; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic proteins

XX from Neisseria genomic sequences. AAA81453 to AAA82414 represent

XX specifically claimed Neisseria meningitidis genomic DNA sequences;

XX AAA81260 to AAA81303 and AAA82520 to AAA82563 represent Neisseria DNA

XX sequences and their corresponding proteins; AAA81254 to AAA81259 and

XX AAA81304 to AAA81321 represent PCR primers used in the isolation of

XX Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent

XX Neisseria meningitidis MenB polynucleotide ORF sequences, which are all

XX used in the exemplification of the present invention. The nucleic acid

XX sequences, protein sequences, and antibodies against them, can be used in

XX the manufacture of a composition. The composition can be used as a

XX medicament (or in the manufacture of a medicament) for treating,

XX preventing or diagnosing infection due to Neisserial bacteria. For

XX example, some of the identified proteins could be components of vaccines

XX against Meningococcus B; against all serotypes; and/or against all

XX pathogenic Neisseriae. Identification of sequences from the bacterium

XX will also facilitate production of biological probes, particularly

XX organism-specific probes. Attempts to make efficacious Meningococcus B

XX vaccines have failed mainly due to antigen tolerance. Multivalent

XX vaccines have also been tried but none have successfully overcome

XX antigenic variability. The provision of further, complete sequences may

XX provide an opportunity to identify secreted or surface exposed proteins

XX that may be presumed targets for the immune system and which are not

XX antigenically variable or at least more conserved than other more

XX variable regions

XX SQ Sequence 23532 BP; 6919 A; 5186 C; 5000 G; 6427 T; 0 U; 0 Other;

XX Query Match 73.6%; Score 20.6; DB 3; Length 23532;

XX Best Local Similarity 85.2%; Pred. No. 96;

XX Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

XX

XX 2 AACACATAACCAAAACGCATATGATT 28

XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

XX 18921 AACCAATAACCAATAAATGCATATCAT 18947

XX

XX RESULT 7

XX AAA81490_14

Continuation (15 of 15) of AAA81490 from base 1400001 (N. meningitidis B full length 9

WP Sequence split into 15 fragments LOCUS AAA81490 Accession AAA81490

WP Fragment Name Begin End

WP AAA81490_00 1 110000

WP AAA81490_01 100001 210000

WP AAA81490_02 200001 310000

WP AAA81490_03 300001 410000

WP AAA81490_04 400001 510000

WP AAA81490_05 500001 610000

WP AAA81490_06 600001 710000

WP AAA81490_07 700001 810000

WP AAA81490_08 800001 910000

WP AAA81490_09 900001 1010000

WP AAA81490_10 1000001 1110000

WP AAA81490_11 1100001 1210000

WP AAA81490_12 1200001 1310000

WP AAA81490_13 1300001 1410000

WP AAA81490_14 1400001 1437668

Query Match 73.6%; Score 20.6; DB 3; Length 37668;

Best Local Similarity 85.2%; Pred. No. 99;

Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AACACATAACCAAAACGCATATGATT 28

DB 9361 AACCAATAACCAATAAATGCATATCAT 9387

RESULT 8

AAA81490_13

Continuation (14 of 15) of AAA81490 from base 1300001 (N. meningitidis B full length 9

WP Sequence split into 15 fragments LOCUS AAA81490 Accession AAA81490

WP Fragment Name Begin End

WP AAA81490_00 1 110000

WP AAA81490_01 100001 210000

WP AAA81490_02 200001 310000

WP AAA81490_03 300001 410000

WP AAA81490_04 400001 510000

WP AAA81490_05 500001 610000

WP AAA81490_06 600001 710000

WP AAA81490_07 700001 810000

WP AAA81490_08 800001 910000

WP AAA81490_09 900001 1010000

WP AAA81490_10 1000001 1110000

WP AAA81490_11 1100001 1210000

WP AAA81490_12 1200001 1310000

WP AAA81490_13 1300001 1410000

WP AAA81490_14 1400001 1437668

Query Match 73.6%; Score 20.6; DB 3; Length 110000;

Best Local Similarity 85.2%; Pred. No. 11e+02;

Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AACACATAACCAAAACGCATATGATT 28

DB 109361 AACCAATAACCAATAAATGCATATCAT 109387

RESULT 9

AAF21610

ID AAF21610 standard; DNA; 349980 BP.

XX

XX AAF21610;

XX

XX 13-MAR-2001 (first entry)

XX

XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:111.

XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;

XX diagnosis; antigen; detection; infection; gene therapy; antibacterial;

XX ds.

XX Neisseria meningitidis.

XX
 PN WO200066791-A1.
 XX
 PD 09-NOV-2000.
 XX
 PF 08-MAR-2000; 2000WO-US005928.
 XX
 PR 30-APR-1999; 98US-0132069P.
 XX
 PR 08-OCT-1999; 98WO-US023573.
 XX
 PR 28-FEB-2000; 2000GB-00004695.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarcellini M, Scarlato V;
 PI Rappuoli R, Frazer CM, Grandi G;
 XX
 DR WPI; 2000-647603/62.
 XX
 PT Neisseria meningitidis B full length genome sequence and open reading
 PT frames are used to detect, treat and prevent Neisserial infections.
 XX
 PS Claim 7; Appendix A; 692pp; English.
 XX
 CC The present invention describes the full length genome of Neisseria
 CC meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613
 CC represent fragments of the NMB genomic sequence, as the sequence was too
 CC long to go in a record on its own it was split into 8 sequences which
 CC overlap each other at the beginning and end of each sequence by 4980 bp
 CC (i.e. the last 4980 bp of AAF21544 is repeated at the beginning of
 CC AAF21607, the last 4980 bp of AAF21607 are repeated at the beginning of
 CC AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins
 CC given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR
 CC primers which are used in the exemplification of the present invention.
 CC The NMB genome and fragments from it have antibacterial activity, and can
 CC be used in vaccines and gene therapy. Neisseria nucleic acids, proteins
 CC and/or antibodies which binds to the proteins can be used in compositions
 CC for treating or preventing infection due to Neisserial bacteria or as a
 CC diagnostic reagent for detecting the presence of Neisserial bacteria or
 CC of antibodies raised to Neisserial bacteria. Computers, computer memory,
 CC computer storage medium or computer databases can be used in a search to
 CC identify open reading frames (ORFs) or coding sequences within the NMB
 CC genome. The DNA sequences provide further opportunities to find antigenic
 CC or immunogenic proteins which are more effective in vaccines than the
 CC outer membrane proteins currently used
 XX
 SQ Sequence 349980 BP; 86771 A; 92803 C; 86340 G; 84066 T; 0 U; 0 Other;
 Query Match 73.6%; Score 20.6; DB 3; Length 349980;
 Best Local Similarity 85.2%; Pred. No. 1.1e+02;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 2 AACACATACCAACCAACGCATATGATT 28
 Db 238973 AACCAATACCAATATGCATATCATT 238999
 RESULT 10
 AAF56298/c
 ID AAF56298 standard; DNA; 1543 BP.
 XX
 AC AAF56298;
 XX
 DT 18-APR-2001 (first entry)
 XX
 DE IS2 element sequence #5.
 XX
 KW Miniature inverted repeat transposable element; MITE; duplication;
 KW regulate; IS2; ds.
 XX
 OS Daucus carota.
 XX

PN WO200105986-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 19-JUL-2000; 2000WO-JP004837.
 XX
 PR 21-JUL-1999; 99JP-00206316.
 XX
 PR 21-JUL-1999; 99JP-00206320.
 XX
 PR 12-JUN-2000; 2000JP-00175825.
 XX
 PA (SAFE-) SAN-EI GEN FFI INC.
 PA (OZEK/) OZEKI Y.
 XX
 PI Ozeki Y, Oyanagi M, Fukuda T, Koda T;
 XX
 DR WPI; 2001-147351/15.
 XX
 PT Novel miniature inverted-repeat transposable element (MITE)-like element
 PT useful for constructing transgene expression cassette for stably creating
 PT genetically engineered organisms capable of expression of transgene.
 XX
 PS Disclosure; Page 103-104; 104pp; English.
 XX
 CC The present invention relates to a miniature inverted-repeat transposable
 CC element (MITE)-like element capable of causing duplication of a target
 CC sequence at the site of its insertion in a genomic gene. The invention is
 CC useful for causing expression of a transgene in a plant. It can also be
 CC used for inducing or regulating the expression in a plant of the gene
 CC introduced
 XX
 SQ Sequence 1543 BP; 569 A; 244 C; 214 G; 516 T; 0 U; 0 Other;
 Query Match 71.4%; Score 20; DB 4; Length 1543;
 Best Local Similarity 82.1%; Pred. No. 1.4e+02;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Qy 1 CCACATACCAACCAACGCATATGATT 28
 Db 360 CCACATACCAACCAACGCATATGATT 333
 RESULT 11
 AAZ67933/c
 ID AAZ67933 standard; DNA; 47 BP.
 XX
 AC AAZ67933;
 XX
 DT 10-SEP-2001 (first entry)
 XX
 DE Human map-related biallelic marker SEQ ID NO:2280.
 XX
 KW Human genome; biallelic marker; high density disequilibrium map;
 KW genomic map; haplotype; phenotype; polymorphic base; genotyping;
 KW haplotyping; hybridisation; identification; characterisation; diagnosis;
 KW single nucleotide polymorphism; SNP; ds.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 PH variation replace(24,A)
 PH /*tag= a
 PH /standard_name= "single nucleotide polymorphism"
 FT
 FT
 FT
 XX
 PN WO9954500-A2.
 XX
 PD 28-OCT-1999.
 XX
 PF 21-APR-1999; 99WO-IB000822.
 XX
 PF 21-APR-1998; 98US-0082614P.
 XX
 PR 23-NOV-1998; 98US-0109732P.
 XX
 PA (GEST) GENSET.


```

XX Cohen D, Blumenfeld M, Chumakov I;
XX WPI; 2000-013267/01.
XX Novel biallelic markers used to construct a high density disequilibrium
XX map of the human genome.
XX Claim 3; Page 717; 2745pp; English.
XX AA265654 to AA269578 represent human biallelic markers from the present
XX invention, which contain a polymorphic base at position 24 of their
XX nucleotide sequences. AA269579 to AA277440 represent amplification
XX primers for the biallelic markers. The biallelic markers of the invention
XX have a variety of uses; they can be used for high density mapping of the
XX human genome, and in complex association studies and haplotyping studies
XX which are useful in determining the genetic basis for disease states.
XX Compositions and methods of the invention can also be useful for the
XX identification of the targets for the development of pharmaceutical
XX agents and diagnostic methods, as well as the characterisation of the
XX differential efficacious responses to and side effects from
XX pharmaceutical agents acting on a disease as well as other treatment.
XX N.B. The SEQ ID NOS 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297 and
XX 3367, are not actually given a sequence in the Sequence Listing from the
XX present invention
XX SQ Sequence 47 BP; 13 A; 3 C; 8 G; 23 T; 0 U; 0 Other;
Query Match 70.0%; Score 19.6; DB 3; Length 47;
Best Local Similarity 84.6%; Pred. No. 1.6e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 3 ACACATAACCAAAACGCATATGAT 28
Db 47 AACATATAACCAAAATGCATATGAT 22
RESULT 12
AAQ67726
ID AAQ67726 standard; DNA; 1491 BP.
AC AAQ67726;
XX 16-OCT-2003 (revised)
XX 25-MAR-2003 (revised)
XX 16-MAR-1995 (first entry)
XX Comamonas testosteroni NI 1 amidase gene.
XX amidase; ammonium adipamate; diammonium adipate; nylon 6.6;
XX 5-cyanovallieramide; 5-cyanovallierate; caprolactam; ss.
XX Comamonas testosteroni; NI 1.
XX Key Location/Qualifiers
XX CDS 127..1383
XX /*tag= a
XX /*product= "amidase"
XX WO9417190-A1.
XX 04-AUG-1994.
XX 21-JAN-1994; 94WO-FR000080.
XX 27-JAN-1993; 93FR-00001062.
XX (RHON ) RHONE POULENC CHIM.
XX Cerbelaud E, Le Coq A, Levyschil S, Petre D, Soubrier F;
XX WPI; 1994-264103/32.
XX P-PSDB; AAR60155.
XX
XX New amidase with greater activity on adipamate than on adipamide - and
XX related DNA and microorganisms producing it, partic used to produce
XX ammonium adipate or 5-cyanovallierate for nylon mfr.
XX Claim 3; Fig 2; 55pp; French.
XX The amidase gene was isolated from a Comamonas testosteroni genomic DNA
XX library using a degenerate probe based on the N-terminal amino acid
XX sequence of the purified enzyme. The amidase is able to hydrolyse amides
XX to carboxylates and has higher activity on ammonium adipamate than on
XX adipamide. The enzyme is particularly useful for converting ammonium
XX adipamate to diammonium adipate for production of nylon 6,6 or for
XX converting 5-cyanovallieramide to 5-cyanovallierate (an intermediate for
XX caprolactam used to make nylon-6). (Updated on 25-MAR-2003 to correct PN
XX field.) (Updated on 25-MAR-2003 to correct PI field.) (Updated on 16-OCT-
XX 2003 to standardise OS field)
XX SQ Sequence 1491 BP; 546 A; 239 C; 262 G; 444 T; 0 U; 0 Other;
Query Match 70.0%; Score 19.6; DB 2; Length 1491;
Best Local Similarity 84.6%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 2 AACACATAACCAAAACGCATATGAT 27
Db 1025 AAAATATAACCAAAACGCATATGAT 1050
RESULT 13
ABL87877/c
ID ABL87877 standard; cDNA; 343 BP.
XX ABL87877;
XX 17-MAY-2002 (first entry)
XX Human ovarian cancer related cDNA clone SEQ ID NO:10855.
XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX Homo sapiens.
XX WO200192581-A2.
XX 06-DEC-2001.
XX 29-MAY-2001; 2001WO-US017756.
XX 26-MAY-2000; 2000US-0207484P.
XX (CORI-) CORIXA CORP.
XX Algate PA, Harlocker SL, Jones R;
XX WPI; 2002-122075/16.
XX Composition for therapy and diagnosis of ovarian cancer comprising
XX polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
XX polypeptide, antibody specific to polypeptide or T cell expressing
XX polypeptide.
XX Claim 1; SEQ ID NO 10855; 489pp; English.
XX The present invention describes a composition (I) comprising: carriers
XX and immunostimulants; and a polypeptide (II) of a ovarian tumour
XX polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
XX from the 10912 nucleotide sequences as given in ABL77023 to ABL7934,
XX (III) encoding (II) having a sequence (S2), a T cell population of (II),
XX or antigen presenting cells that express (II). (I) has cytostatic
XX activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
XX detecting ovarian cancer in a patient's biological sample preferably
XX serum or ovarian tissue. The method comprises contacting a biological

```

CC sample from a patient with (IV), detecting the amount of polynucleotide
CC hybridising to (IV) and comparing the amount to a predetermined cutoff
CC value and thereby detecting ovarian cancer in the patient, where the
CC amount of polynucleotide hybridising to (IV) is detected preferably by
CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC useful for stimulating and/or expanding T cells specific for an ovarian
CC tumour protein comprising contacting T cells with (III) or (II). (III) is
CC useful in design and preparation of ribozyme molecules for inhibiting
CC expression of the tumour polypeptides and proteins in tumour cells; and
CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC library using well known techniques
XX
SQ Sequence 343 BP; 117 A; 52 C; 50 G; 124 T; 0 U; 0 Other;

Query Match 68.6%; Score 19.2; DB 6; Length 343;
Best Local Similarity 87.5%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ACATACCAACCAACGCATATGATT 28
DB 103 ACATACCAACCAACGCATATGATT 80

RESULT 14
ABL32742/c
ID ABL32742 standard; DNA; 8087 BP.

AC ABL32742;

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 715.

XX Human; immune system disease; cytosine methylation; asthenic; antiarteriosclerotic; antianemic; cytosatic; neutropenic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
ds.

XX Homo sapiens.

OS WO200200928-A2.

PN 03-JAN-2002.

PD 02-JUL-2001; 2001WO-EP007537.

PP 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPIC-) EPICENOMICS AG.

PA Olek A, Piepenbrock C, Berlin K;

PI WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.

PS Claim 1; SEQ ID NO 715; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention

XX SQ Sequence 8087 BP; 2081 A; 179 C; 1817 G; 4010 T; 0 U; 0 Other;
Query Match 68.6%; Score 19.2; DB 6; Length 8087;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAACACATACCAACCAACGCATAT 24
DB 675 CAACACATACCAACCAACGCATAT 652

RESULT 15

AAH52751/c

ID AAH52751 standard; DNA; 945 BP.

XX AC AAH52751;

XX 03-SEP-2001 (first entry)

XX S. epidermidis open reading frame nucleotide sequence SEQ ID NO:895.

DE Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
KW endocarditis; ds.

XX Staphylococcus epidermidis.

OS WO200134809-A2.

PN 17-MAY-2001.

PP 09-NOV-2000; 2000WO-US030782.

PR 09-NOV-1999; 99US-0164258P.

XX (GLAX) GLAXO GROUP LTD.

XX Kimmerly WJ;

XX WPI; 2001-316495/33.

DR P-PSDB; AXG81901.

XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis.

PS Claim 8; Page 267; 2188pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AXG81454 to AXG83120, from Staphylococcus epidermidis. (I)
CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acids (I) may be used to produce the S.
CC epidermidis polypeptides (II) via the production of vectors containing
CC them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC represent oligonucleotide sequences and primers which are used in the
CC amplification of the present invention. N.B. The present invention
CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC for SEQ ID NO:4455 to 4464

XX SQ Sequence 945 BP; 390 A; 102 C; 184 G; 269 T; 0 U; 0 Other;

Query Match 67.9%; Score 19; DB 4; Length 945;
Best Local Similarity 81.5%; Pred. No. 3.4e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 CAACACATAACCAAAAACGCATATGAT 27
Db 493 CAACACATAATAAAAACATGCATATGTT 467

Search completed: May 26, 2004, 17:50:01
Job time : 201.949 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 16:18:00 ; Search time 44.6325 Seconds
(without alignments)
348.146 Million cell updates/sec

Title: US-10-676-299-4

Perfect score: 28

Sequence: 1 caacacataacacaaacgcataatgatt 28

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 632709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/2/ina/5A COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.2	68.6	47	US-09-422-978-2280	Sequence 2280, Ap
C 2	19	67.9	2481	US-09-328-352-3389	Sequence 3389, Ap
C 3	19	67.9	9058	US-08-913-0114A-9	Sequence 9, Appli
C 4	19	67.9	9058	US-09-653-285-9	Sequence 9, Appli
C 5	18.8	67.1	1830121	US-09-557-884-1	Sequence 1, Appli
C 6	18.8	67.1	1830121	US-09-643-990A-1	Sequence 1, Appli
C 7	18.6	66.4	459	US-09-328-352-300	Sequence 300, Ap
C 8	18.4	65.7	567	US-09-107-532A-1957	Sequence 1957, Ap
C 9	18	64.3	19513	US-10-204-708-39	Sequence 39, Appli
C 10	17.8	63.6	1026	US-09-134-001C-1995	Sequence 1995, Ap
C 11	17.8	62.9	202001	US-09-734-674-3	Sequence 3, Appli
C 12	17.6	62.9	1442	US-09-227-357-119	Sequence 119, Ap
C 13	17.6	62.9	4868	US-08-139-937-12	Sequence 12, Appl
C 14	17.6	62.9	4868	PCT-US93-11310-12	Sequence 12, Appl
C 15	17.6	62.9	8789	US-08-328-254-5	Sequence 5, Appli
C 16	17.4	62.1	1253	US-09-596-594-538	Sequence 538, Ap
C 17	17.4	62.1	2061	US-09-107-532A-1464	Sequence 1464, Ap
C 18	17.4	62.1	2311	US-09-349-740A-16	Sequence 16, Appl
C 19	17.4	62.1	3001	US-09-539-333D-178	Sequence 178, Ap
C 20	17.4	62.1	4500	US-09-349-740A-1	Sequence 1, Appli
C 21	17.4	62.1	53588	US-09-873-404-3	Sequence 3, Appli
C 22	17.4	62.1	75395	US-09-984-890-3	Sequence 3, Appli
C 23	17.4	62.1	640681	US-09-790-988-1	Sequence 1, Appli
C 24	17.4	62.1	786431	US-09-751-389-3	Sequence 3, Appli
C 25	17.2	61.4	1070	US-09-718-841-3	Sequence 3, Appli
C 26	17.2	61.4	1070	US-09-718-810-3	Sequence 3, Appli
C 27	17.2	61.4	1070	US-09-718-854-3	Sequence 3, Appli

C 28 17.2 61.4 1689 4 US-09-718-841-1 Sequence 1, Appli
C 29 17.2 61.4 1689 4 US-09-718-810-1 Sequence 1, Appli
C 30 17.2 61.4 1689 4 US-09-718-854-1 Sequence 1, Appli
C 31 17.2 61.4 2418 4 US-09-620-312D-158 Sequence 158, App
C 32 17.2 61.4 580073 4 US-08-545-528D-1 Sequence 1, Appli
C 33 17.2 61.4 1664976 4 US-08-316-421B-1 Sequence 1, Appli
C 34 17 60.7 653 4 US-09-702-705-313 Sequence 313, App
C 35 17 60.7 653 4 US-09-736-457-313 Sequence 313, App
C 36 17 60.7 653 4 US-09-614-124B-313 Sequence 313, App
C 37 17 60.7 653 4 US-09-671-325-313 Sequence 313, App
C 38 17 60.7 653 4 US-09-589-184-313 Sequence 313, App
C 39 17 60.7 2800 3 US-09-188-930-260 Sequence 260, App
C 40 17 60.7 2800 4 US-09-312-283C-260 Sequence 260, App
C 41 17 60.7 19124 2 US-08-487-826B-13 Sequence 13, Appli
C 42 17 60.7 580073 4 US-08-545-528D-1 Sequence 1, Appli
C 43 16.8 60.0 271 4 US-09-313-294A-6454 Sequence 6454, Ap
C 44 16.8 60.0 483 4 US-09-325-932A-5 Sequence 5, Appli
C 45 16.8 60.0 774 3 US-08-221-767-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-422-978-2280/c

; Sequence 2280, Application US/09422978

; Patent No. 6537751

; GENERAL INFORMATION:

; APPLICANT: Cohen, Daniel

; APPLICANT: Blumenfeld, Marta

; APPLICANT: Chumakov, Ilya

; TITLE OF INVENTION: Blallelic markers for use in constructing a high density...

; FILE REFERENCE: GENSET.020CPI

; CURRENT APPLICATION NUMBER: US/09/422,978

; CURRENT FILING DATE: 1999-10-20

; EARLIER APPLICATION NUMBER: US 09/298,850

; EARLIER FILING DATE: 1999-04-21

; EARLIER APPLICATION NUMBER: US 60/109,732

; EARLIER FILING DATE: 1998-11-23

; EARLIER APPLICATION NUMBER: US 60/082,614

; EARLIER FILING DATE: 1998-04-21

; NUMBER OF SEQ ID NOS: 11796

; SEQ ID NO 2280

; LENGTH: 47

; TYPE: DNA

; ORGANISM: Homo Sapiens

; FEATURE:

; NAME/KEY: allele

; LOCATION: 24

; OTHER INFORMATION: 99-10146-202 : polymorphic base T or A

US-09-422-978-2280

Query Match 68.6%; Score 19.2; DB 4; Length 47;

Best Local Similarity 80.8%; Pred. No. 15;

Matches 21; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 ACACATACCAAAACGCATATGATT 28

||||| ||||| ||||| ||||| :||

DB 47 AACATTAACAAAATGCATATATTT 22

RESULT 2

US-09-328-352-3389/c

; Sequence 3389, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breston et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

US-09-653-285-9/c
; Sequence 9, Application US/09653285
; Patent No. 6530090
; GENERAL INFORMATION:
; APPLICANT: Nishi, Kazunori
; HIKICHI, YUKIKO
; SHINTANI, YASUSHI
; TITLE OF INVENTION: NOVEL FAS LIGAND-LIKE PROTEIN, ITS
; PRODUCTION AND USE
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David G. Conlin, Esq.
; DIKE, BRONSTEIN, ROBERTS & USHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston,
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/653,285
; FILING DATE: 31-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/913,014
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: David G. Conlin
; REGISTRATION NUMBER: 27,026
; REFERENCE/DOCKET NUMBER: 342/47694
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9058
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Linear
; MOLECULE TYPE: genomic DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-653-285-9
Query Match 67.9%; Score 19; DB 4; Length 9058;
Best Local Similarity 81.5%; Pred. No. 37;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 2 AACACATAACCAAAAGCGCATATGATT 28
Db 8604 AACAAATAACAAAGCACAATGATT 8578
RESULT 5
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: the Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA

SEQ ID NO 3389
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3389
Query Match 67.9%; Score 19; DB 4; Length 2481;
Best Local Similarity 81.5%; Pred. No. 31;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 2 AACACATAACCAAAAGCGCATATGATT 23
Db 1941 AACCCATAAACCAAGCGCATATCATTT 1915
RESULT 3
US-08-913-014A-9/c
; Sequence 9, Application US/08913014A
; Patent No. 6235878
; GENERAL INFORMATION:
; APPLICANT: Nishi, Kazunori
; APPLICANT: HIKICHI, YUKIKO
; APPLICANT: SHINTANI, YASUSHI
; TITLE OF INVENTION: NOVEL FAS LIGAND-LIKE PROTEIN, ITS
; PRODUCTION AND USE
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David G. Conlin, Esq.
; DIKE, BRONSTEIN, ROBERTS & USHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston,
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,014A
; FILING DATE: 04-SEP-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP97/02480
; FILING DATE: July 17, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: David G. Conlin
; REGISTRATION NUMBER: 27,026
; REFERENCE/DOCKET NUMBER: 342/47694
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9058
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Linear
; MOLECULE TYPE: genomic DNA
US-08-913-014A-9
Query Match 67.9%; Score 19; DB 3; Length 9058;
Best Local Similarity 81.5%; Pred. No. 37;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 2 AACACATAACCAAAAGCGCATATGATT 28
Db 8604 AACAAATAACAAAGCACAATGATT 8578
RESULT 4

```

;
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 67.1%; Score 18.8; DB 4; Length 1830121;
Best Local Similarity 90.9%; Pred. No. 81;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 ATAACCAAAAGCGCATATGATT 28
|||||
Db 1238890 ATAACCAAAAGCGCATATGTT 1238911

RESULT 6
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21

```

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 67.1%; Score 18.8; DB 4; Length 1830121;
Best Local Similarity 90.9%; Pred. No. 81;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 ATAACCAAAAGCGCATATGATT 28
|||||
Db 1238990 ATAACCAAAAGCGCATATGTT 1238911

RESULT 7
; Sequence 300, Application US/09328352
; Patent No. 6562358
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 300
; LENGTH: 459
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-300

Query Match 66.4%; Score 18.5; DB 4; Length 459;
Best Local Similarity 84.0%; Pred. No. 36;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CACATAACCAAAAGCGCATATGATT 28
|||||
Db 98 CACATAACCAAAAGCGCACTGAT 74

RESULT 8
US-09-107-532A-1957/c
; Sequence 1957, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A. Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIORITY INFORMATION:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 1957:
SEQUENCE CHARACTERISTICS:
LENGTH: 567 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...567
SEQUENCE DESCRIPTION: SEQ ID NO: 1957:
US-09-107-532A-1957

```

Query Match 65.7%; Score 18.4; DB 4; Length 567;
 Best Local Similarity 78.6%; Pred. No. 45;
 Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CAACATATACCAAAACGCATATGATT 28
 |||||
 Db 120 CAATCATATCCATCCGCAATATGATT 93

```

RESULT 9
US-10-204-708-39/c
Sequence 39, Application US/10204708
Patent No. 667731
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/204,708
CURRENT FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 39
LENGTH: 19513
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-39

```

Query Match 64.3%; Score 18; DB 4; Length 19513;
 Best Local Similarity 80.8%; Pred. No. 1.1e+02;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AACACATACCAAAACGCATATGAT 27
 |||||
 Db 18377 AAAACATATCAAAACGCATATAT 18352

```

RESULT 10
US-09-134-001C-1995
Sequence 1995, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 1995
LENGTH: 1026
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1995

```

Query Match 63.6%; Score 17.8; DB 4; Length 1026;
 Best Local Similarity 90.5%; Pred. No. 88;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 8 TAACCAAAACGCATATGATT 28
 |||||
 Db 602 TACTAAAAAGCATATGATT 522

```

RESULT 11
US-09-734-674-3/c
Sequence 3, Application US/09734674
Patent No. 6498022
GENERAL INFORMATION:
APPLICANT: WEI, Ming-Hui et al
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: CL001018
CURRENT APPLICATION NUMBER: US/09/734,674
CURRENT FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 202001
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(202001)
OTHER INFORMATION: n = A,T,C or G
US-09-734-674-3

```

Query Match 63.6%; Score 17.8; DB 4; Length 202001;
 Best Local Similarity 90.5%; Pred. No. 1.8e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 6 CATAACCAAAACGCATATGCA 26
 |||||
 Db 175136 CATAACCAAAAGCATAGGA 175116

```

; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 119
; LENGTH: 1442
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1377)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1419)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-227-357-119
;
Query Match 62.9%; Score 17.6; DB 4; Length 1442;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0
;
QY 5 ACATAACCAAAACGCATATGATT 28
||| |||| |||| |||| ||||
DB 774 ACACACCAATAACAAATATGATT 751
;
RESULT 13
US-08-139-937-12/C
; Sequence 12, Application US/08139937
; Patent No. 5821070
; GENERAL INFORMATION:
; APPLICANT: LEE, WEN-HWA
; APPLICANT: SHAN, BEI
; TITLE OF INVENTION: CELLULAR GENES ENCODING
; TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/139,937
; FILING DATE: 20-OCT-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/979,156
; FILING DATE: 20-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHRYN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 9370
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4868 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-139-937-12

```


Query Match 62.9%; Score 17.6; DB 1; Length 4868;

Best Local Similarity 83.3%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AACACATACCAAAACGCATATG 25

Db 4689 AACACATACCCACACACATATG 4666

RESULT 14

PCT-US93-11310-12/c

; Sequence 12, Application PC/TUS9311310

; GENERAL INFORMATION:

; APPLICANT: BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM

; TITLE OF INVENTION: CELLULAR GENES ENCODING

; TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CAMPBELL AND FLORES

; STREET: 4370 LA JOLLA VILLAGE DRIVE

; CITY: SAN DIEGO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US93/11310

; FILING DATE: 19-NOV-1993

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: CAMPBELL, CATHRYN

; REGISTRATION NUMBER: 31,815

; REFERENCE/DOCKET NUMBER: FP-CJ 9790

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 619-535-9001

; TELEFAX: 619-535-8949

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4868 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

PCT-US93-11310-12

Query Match 62.9%; Score 17.6; DB 5; Length 4868;

Best Local Similarity 83.3%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AACACATACCAAAACGCATATG 25

Db 4689 AACACATACCCACACACATATG 4666

RESULT 15

US-08-328-254-5/c

; Sequence 5, Application US/08328254

; Patent No. 5710022

; GENERAL INFORMATION:

; APPLICANT: Zhu, Xueliang

; APPLICANT: Lee, Wen-Hwa

; TITLE OF INVENTION: A No. 5710022e1 Nuclear Mitotic Phosphoprotein

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: USA

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/328,254

; FILING DATE: 24-OCT-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/141,239

; FILING DATE: 22-OCT-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Campbell, Cathryn A.

; REGISTRATION NUMBER: 31,815

; REFERENCE/DOCKET NUMBER: P-CJ 1191

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 535-9001

; TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 8789 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 544..7990

US-08-328-254-5

Query Match 62.9%; Score 17.6; DB 1; Length 8789;

Best Local Similarity 83.3%; Pred. No. 1.4e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AACACATACCAAAACGCATATG 25

Db 8590 AACACATACCCACACACATATG 8567

Search completed: May 26, 2004, 17:56:36

Job time : 52.6325 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 17:50:29 ; Search time 574.838 Seconds
(without alignments)
221.574 Million cell updates/sec

Title: US-10-676-299-4

Perfect score: 28
Sequence: 1 CAACACATAACCAAAACGCATATGATT 28

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2960401 seqs, 2274450654 residues

Total number of hits satisfying chosen parameters: 5920802

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	100.0	28	15	US-10-222-952A-4
2	28	100.0	30	15	US-10-222-952A-3
3	21.6	77.1	79860	13	US-10-087-192-412
4	21	75.0	21	15	US-10-222-952A-8
5	21	75.0	23	15	US-10-222-952A-7
6	20.6	73.6	6391	15	US-10-240-453-11
7	20	71.4	741	13	US-10-037-632-166036
8	20	71.4	741	16	US-10-027-632-166036
9	20	71.4	248436	13	US-10-087-192-2014
10	19.8	70.7	251364	15	US-10-175-523-58
11	19.8	70.7	251364	15	US-10-175-523-61
12	19.8	70.7	251364	15	US-10-175-523-79
13	19.2	68.6	47	16	US-10-349-143-2280
14	19.2	68.6	343	9	US-09-867-701-10855

C 15	19.2	68.6	8087	15	US-10-311-455-715	Sequence 715, App
C 16	19.2	68.6	106378	13	US-10-087-192-1624	Sequence 1624, App
C 17	19	67.9	520	13	US-10-027-632-204483	Sequence 204483, App
C 18	19	67.9	520	16	US-10-027-632-204483	Sequence 204483, App
C 19	19	67.9	577	13	US-10-425-114-19376	Sequence 19376, A
C 20	19	67.9	684	13	US-10-027-632-129002	Sequence 129002, A
C 21	19	67.9	684	16	US-10-027-632-129002	Sequence 129002, A
C 22	19	67.9	2222	13	US-10-027-632-260729	Sequence 260729, A
C 23	19	67.9	2222	16	US-10-027-632-260729	Sequence 260729, A
C 24	19	67.9	9058	13	US-10-468-161-9	Sequence 9, Appli
C 25	19	67.9	12103	10	US-09-764-891-7733	Sequence 7733, App
C 26	19	67.9	17534	17	US-10-257-166-107	Sequence 107, App
C 27	19	67.9	34769	13	US-10-221-714A-501	Sequence 501, App
C 28	18.8	67.1	1830121	15	US-10-329-960-1	Sequence 1, Appli
C 29	18.8	67.1	1830121	16	US-10-329-960-1	Sequence 1, Appli
C 30	18.6	66.4	607	13	US-10-027-632-75585	Sequence 75585, A
C 31	18.6	66.4	607	13	US-10-027-632-75586	Sequence 75586, A
C 32	18.6	66.4	607	13	US-10-027-632-76240	Sequence 76240, A
C 33	18.6	66.4	607	13	US-10-027-632-76241	Sequence 76241, A
C 34	18.6	66.4	607	13	US-10-027-632-320570	Sequence 320570, A
C 35	18.6	66.4	607	13	US-10-027-632-320571	Sequence 320571, A
C 36	18.6	66.4	607	16	US-10-027-632-75585	Sequence 75585, A
C 37	18.6	66.4	607	16	US-10-027-632-75586	Sequence 75586, A
C 38	18.6	66.4	607	16	US-10-027-632-76240	Sequence 76240, A
C 39	18.6	66.4	607	16	US-10-027-632-76241	Sequence 76241, A
C 40	18.6	66.4	607	16	US-10-027-632-320570	Sequence 320570, A
C 41	18.6	66.4	607	16	US-10-027-632-320571	Sequence 320571, A
C 42	18.6	66.4	617	13	US-10-027-632-47688	Sequence 47688, A
C 43	18.6	66.4	617	13	US-10-027-632-47689	Sequence 47689, A
C 44	18.6	66.4	617	16	US-10-027-632-47688	Sequence 47688, A
C 45	18.6	66.4	617	16	US-10-027-632-47689	Sequence 47689, A

ALIGNMENTS

RESULT 1

US-10-222-952A-4
; Sequence 4, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regensis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/IL443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 28
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PLASL1B bottom, long oligo sequence
US-10-222-952A-4

Query Match 100.0%; Score 28; DB 15; Length 28;

Best Local Similarity 100.0%; Pred. No. 0.11; 0; Indels 0; Gaps 0; Matches 28; Conservative 0; Mismatches 0;

Qy 1 CAACACATAACCAAAACGCATATGATT 28

Db 1 CAACACATAACCAAAACGCATATGATT 28

RESULT 2

US-10-222-952A-3/c

; Sequence 3, Application US/10222952A

; Publication No. US20030096275A1

; GENERAL INFORMATION:

```
; APPLICANT: Regensis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/1L443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; PRIOR FILING DATE: 2002-08-15
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PLASLIT biotinylated top, long oligo sequence; biotinylated
; OTHER INFORMATION: nucleotide at position 1
US-10-222-952A-3

Query Match      100.0%; Score 28; DB 15; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAACACATAACCAAAAACGCATATGATT 28
   |||||
Db 30 CAACACATAACCAAAAACGCATATGATT 3

RESULT 3
US-10-087-192-412
; Sequence 412, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR FILING DATE: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 412
; LENGTH: 79860
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(79860)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-412
```

```
Query Match      77.1%; Score 21.6; DB 13; Length 79860;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CAACACATAACCAAAAACGCATATGATT 28
   |||||
Db 65272 CAACATACCAAAAACGCATATGATT 65299
```

```
RESULT 4
US-10-222-952A-8
; Sequence 8, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regensis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
```

```
; FILE REFERENCE: 4107/1L443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PLASS1B bottom, short oligo sequence
US-10-222-952A-8
```

```
Query Match      75.0%; Score 21; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 8 TAACCAAAAACGCATATGATT 28
   |||||
Db 1 TAACCAAAAACGCATATGATT 21
```

```
RESULT 5
US-10-222-952A-7/c
; Sequence 7, Application US/10222952A
; Publication No. US20030096275A1
; GENERAL INFORMATION:
; APPLICANT: Regensis
; APPLICANT: Laing, Lance
; TITLE OF INVENTION: BIOSENSOR FOR SMALL MOLECULE ANALYTES
; FILE REFERENCE: 4107/1L443-US1
; CURRENT APPLICATION NUMBER: US/10/222,952A
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/313,714
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 23
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PLASS1T top, short biotinylated oligo sequence; biotinylated
; OTHER INFORMATION: nucleotide at position 1
US-10-222-952A-7
```

```
Query Match      75.0%; Score 21; DB 15; Length 23;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 8 TAACCAAAAACGCATATGATT 28
   |||||
Db 23 TAACCAAAAACGCATATGATT 3
```

```
RESULT 6
US-10-240-453-11/c
; Sequence 11, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
; TITLE OF INVENTION: With DNA Transcription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240,453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
```

```
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032525.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043825.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 11
; LENGTH: 6391
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-11

Query Match      73.6%; Score 20.6; DB 15; Length 6391;
Best Local Similarity 85.2%; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2  AACACATACCAAAACGCGATGATT 28
Db      1027 AACACATCCAAATACACATATT 1001

RESULT 7
US-10-027-632-166036
; Sequence 166036, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 166036
; LENGTH: 741
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-166036

Query Match      71.4%; Score 20; DB 16; Length 741;
Best Local Similarity 82.1%; Pred. No. 2.8e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1  CAACACATACCAAAACGCGATGATT 28
Db      275  CAAACAAACAAACAAACTCATGTGATT 302

RESULT 8
US-10-027-632-166036
; Sequence 166036, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 166036
; LENGTH: 741
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-166036

Query Match      71.4%; Score 20; DB 13; Length 741;
Best Local Similarity 82.1%; Pred. No. 2.8e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1  CAACACATACCAAAACGCGATGATT 28
Db      275  CAAACAAACAAACAAACTCATGTGATT 302

RESULT 9
US-10-087-192-2014/c
; Sequence 2014, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2014
; LENGTH: 248436
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(248436)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-2014

Query Match      71.4%; Score 20; DB 13; Length 248436;
Best Local Similarity 82.1%; Pred. No. 7.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1  CAACACATACCAAAACGCGATGATT 28
Db      76044  CAAACAAACAAACAAACTCATGTGATT 76017
```

```
RESULT 10
US-10-175-523-58/c
; Sequence 58, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Palfreyman, Michael
; APPLICANT: Rajan, Prithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/1J795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58
; LENGTH: 251364
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-175-523-58

Query Match      70.7%; Score 19.8; DB 15; Length 251364;
Best Local Similarity 91.3%; Pred. No. 9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CAACACATAACCAAAACGCATA 23
Db      84700 CAACACATAACTTAAACGCATA 84678

RESULT 11
US-10-175-523-61/c
; Sequence 61, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Palfreyman, Michael
; APPLICANT: Rajan, Prithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/1J795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59
; LENGTH: 251364
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-175-523-61

Query Match      70.7%; Score 19.8; DB 15; Length 251364;
Best Local Similarity 91.3%; Pred. No. 9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CAACACATAACCAAAACGCATA 23
Db      84700 CAACACATAACTTAAACGCATA 84678

RESULT 12
US-10-175-523-79/c
; Sequence 79, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Palfreyman, Michael
; APPLICANT: Rajan, Prithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/1J795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 79
; LENGTH: 251364
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-175-523-79

Query Match      70.7%; Score 19.8; DB 15; Length 251364;
Best Local Similarity 91.3%; Pred. No. 9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CAACACATAACCAAAACGCATA 23
Db      84700 CAACACATAACTTAAACGCATA 84678

RESULT 13
US-10-349-143-2280/c
; Sequence 2280, Application US/10349143
; Publication No. US20040005584A1
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: Biallelic markers for use in constructing a high density...
```

```
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 251364
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-175-523-61

Query Match      70.7%; Score 19.8; DB 15; Length 251364;
Best Local Similarity 91.3%; Pred. No. 9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CAACACATAACCAAAACGCATA 23
Db      84700 CAACACATAACTTAAACGCATA 84678

RESULT 12
US-10-175-523-79/c
; Sequence 79, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Palfreyman, Michael
; APPLICANT: Rajan, Prithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/1J795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 79
; LENGTH: 251364
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-175-523-79

Query Match      70.7%; Score 19.8; DB 15; Length 251364;
Best Local Similarity 91.3%; Pred. No. 9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CAACACATAACCAAAACGCATA 23
Db      84700 CAACACATAACTTAAACGCATA 84678

RESULT 13
US-10-349-143-2280/c
; Sequence 2280, Application US/10349143
; Publication No. US20040005584A1
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: Biallelic markers for use in constructing a high density...
```

FILE REFERENCE: GENSET.020CP1
CURRENT APPLICATION NUMBER: US/10/349,143
CURRENT FILING DATE: 2003-01-21
PRIOR APPLICATION NUMBER: US/09/422,978
PRIOR FILING DATE: 1999-10-20
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 09/298,850
PRIOR FILING DATE: EARLIER FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/109,732
PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-23
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/082,614
PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-21
NUMBER OF SEQ ID NOS: 11796
SEQ ID NO 2280
LENGTH: 47
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KRY: allele
LOCATION: 24
OTHER INFORMATION: 99-10146-202 : polymorphic base T or A
US-10-349-143-2280

Query Match 68.6%; Score 19.2; DB 16; Length 47;
Best Local Similarity 80.8%; Pred. No. 3.5e+02;
Matches 21; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 ACACATAACCAAAAGCGCATATGATT 28
|||||
Db 47 AAACATAAACCAAAAGCGCATATGATT 22
|||||

RESULT 14

US-09-867-701-10855/c
Sequence 10855, Application US/09867701
Patent No. US20020132237A1
GENERAL INFORMATION:
APPLICANT: Jones, Robert
APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.497
CURRENT APPLICATION NUMBER: US/09/867,701
CURRENT FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 10912
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10855
LENGTH: 343
TYPE: DNA
ORGANISM: Homo sapien
US-09-867-701-10855

Query Match 68.6%; Score 19.2; DB 9; Length 343;
Best Local Similarity 87.5%; Pred. No. 5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ACATAACCAAAAGCGCATATGATT 28
|||||
Db 103 ACATAACTAAAGGCACTCTGATT 80
|||||

RESULT 15

US-10-311-455-715/c
Sequence 715, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: FIEPENBROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Methylation Status of Cytosine
TITLE OF INVENTION: cytosine methylation
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455

CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-03-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 715
LENGTH: 8087
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-715

Query Match 68.6%; Score 19.2; DB 15; Length 8087;
Best Local Similarity 87.5%; Pred. No. 8.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAACACATAACCAAAAGCGCATAT 24
|||||
Db 675 CAACACTTAATATAAAAGCGCATAT 652
|||||

Search completed: May 27, 2004, 06:16:17
Job time : 578.838 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 15:22:28 ; Search time 299.923 Seconds

(without alignments)
594.900 Million cell updates/sec

Title: US-10-676-299-5

Perfect score: 42

Sequence: 1 ctgcacttacacattgta.....tcatatgttttgactta 42

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747736

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseqn_29Jan04.*

1: geneseqn1980s.*
2: geneseqn1990s.*
3: geneseqn2000s.*
4: geneseqn2001as.*
5: geneseqn2001bs.*
6: geneseqn2002s.*
7: geneseqn2003as.*
8: geneseqn2003bs.*
9: geneseqn2003cs.*
10: geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	42	8	ACD28586 E. coli A
2	42	100.0	401	7	ACA15469 Prokaryot
3	42	100.0	2182	5	AAS82555 DNA encod
4	40	95.2	40	8	ACD28587 E. coli A
5	25	59.5	25	8	ACD28591 E. coli A
6	25	59.5	25	8	ACD28590 E. coli A
7	23.4	55.7	29956	8	ADA02963 Mouse Lck
8	23.4	55.7	29956	9	ADB72701 Mouse Lck
9	23.4	55.7	29956	9	ADC85443 Mouse Lck
10	22.8	54.3	396	6	ABN25854 Human ORF
11	22.6	53.8	3503	4	ABL21302 Drosophil
12	22.6	53.8	6872	4	ABL18112 Drosophil
13	22.4	53.3	3109	4	ABU05632 Drosophil
14	22.2	52.9	1773	6	ABQ71197 Listeria
15	22.2	52.9	8148	6	ABQ71020 Listeria
16	22	52.4	3647	4	ABL11816 Drosophil
17	21.8	51.9	1146	7	ACA49596 Prokaryot
18	21.8	51.9	1166	2	AAQ04704 USP-Promo
19	21.8	51.9	2000	6	ABZ17501 Arabidops
20	21.8	51.9	3431	4	ABL05396 Drosophil
21	21.8	51.9	3741	4	ABL05790 Drosophil
22	21.8	51.9	3856	4	ABL05792 Drosophil
23	21.8	51.9	5082	4	ABL04626 Drosophil

24	21.8	51.9	7820	4	ABL29332 Drosophil
25	21.8	51.9	8004	4	ABL29346 Drosophil
26	21.8	51.9	8033	4	ABL19178 Drosophil
27	21.8	51.9	19082	6	ABL32626 Human imm
28	21.6	51.4	15667	6	ABL34146 Human imm
29	21.6	51.4	15933	4	ABL30408 Drosophil
30	21.6	51.4	17929	9	ADC86000 Human GPC
31	21.6	51.4	41104	6	AAD36260 Human G-p
32	21.4	51.0	534	5	ABV51050 Human pro
33	21.4	51.0	660	6	ABQ21634 Oligonucl
34	21.4	51.0	660	6	ABQ21635 Oligonucl
35	21.4	51.0	1719	8	ABZ32449 Canida a
36	21.4	51.0	1733	3	AAC45407 Arabidops
37	21.4	51.0	1848	3	AAC45408 Arabidops
38	21.2	50.5	405	7	ABX62578 Arabidops
39	21.2	50.5	1260	3	AZ45412 D-ribulos
40	21.2	50.5	1287	7	ACF70667 Photorhab
41	21.2	50.5	2000	7	ADA73440 Rice gene
42	21.2	50.5	4282	5	AAS86938 DNA encod
43	21.2	50.5	7823	4	AAS45490 Chemical
44	21.2	50.5	7823	6	ABL34061 Human imm
45	21.2	50.5	7823	6	ABK31493 Signal tr

ALIGNMENTS

RESULT 1

ACD28586
ID ACD28586 standard; DNA; 42 BP.
XX
AC ACD28586;
XX
DT 10-OCT-2003 (first entry)
XX
DE E. coli Arsr binding oligonucleotide CHROMLIT.
XX
KW Arsr; arsenic resistance operon; biosensor; ss; arsenic.
XX
OS Escherichia coli.
XX
FN US2003096275-A1.
XX
PD 22-MAY-2003.
XX
PF 15-AUG-2002; 2002US-00222952.
XX
PR 20-AUG-2001; 2001US-0313714P.
XX
PA (LAIN/) LAING L G.
XX
PI Laing LG;
XX
DR WPI; 2003-576876/54.
XX
PT New system comprising isolated protein and nucleic acid, and a detection system that indicates a change in binding of the protein to the nucleic acid in the presence of the analyte, useful for detecting the presence of an analyte in a sample.
XX
PS Claim 35; Page 15; 36pp; English.

The invention relates to a new system (biosensor) for detecting the presence of analyte in a sample comprising: (1) an isolated protein that specifically binds the analyte; (2) an isolated nucleic acid containing a specific binding sequence that is bound specifically by the protein; and (3) a detection system that indicates a change in binding of the protein to the nucleic acid in the presence of the analyte. Also included are a biosensor device for detecting the presence of an analyte in a sample, detecting the presence of an analyte in a sample, and an Arsr (encoded by part of the arsenic resistance operon of E. coli) protein comprising an amino acid sequence that is at least 90% identical to amino acids 1-97 of the Arsr sequence appearing as ABU63440 binding to a nucleic acid

CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the top strand of an oligonucleotide which binds to a
 CC chromosomally expressed ArsR protein and is used in the biosensor of the
 CC invention

XX Sequence 42 BP; 11 A; 8 C; 5 G; 18 T; 0 U; 0 Other;
 SQ Query Match 100.0%; Score 42; DB 8; Length 42;
 Best Local Similarity 100.0%; Pred. No. 8.1e-06;
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCACCTTACACATTCGTTAAGTCATATATATGTTTGACTTA 42
 DB 1 CTGCACCTTACACATTCGTTAAGTCATATATATGTTTGACTTA 42

RESULT 2
 ACA15469/c
 ID ACA15469 standard; DNA; 401 BP.
 XX ACA15469;
 AC ACA15469;
 XX 27-OCT-2003 (revised)
 DT 19-JUN-2003 (first entry)
 XX Prokaryotic essential gene antisense oligonucleotide #3339.
 XX Antisense; ss; prokaryotic essential gene; cell proliferation;
 KW drug design.
 XX Archaea.
 OS WO200277183-A2.
 PN WO200277183-A2.
 PD 03-OCT-2002.
 XX 21-MAR-2002; 2002WO-US009107.
 XX 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX {ELIT-} ELITRA PHARM INC.
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX Claim 1; SEQ ID NO 3339; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the 6213
 CC antisense sequences of the invention. Note: the sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to
 CC standardise OS field)

SQ Sequence 401 BP; 113 A; 68 C; 101 G; 119 T; 0 U; 0 Other;
 Query Match 100.0%; Score 42; DB 7; Length 401;
 Best Local Similarity 100.0%; Pred. No. 1e-05;
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCACCTTACACATTCGTTAAGTCATATATATGTTTGACTTA 42
 DB 145 CTGCACCTTACACATTCGTTAAGTCATATATATGTTTGACTTA 104

RESULT 3
 AAS82555
 ID AAS82555 standard; cDNA; 2182 BP.
 XX AAS82555;
 AC AAS82555;
 XX 13-FEB-2002 (first entry)
 DT DNA encoding novel human diagnostic protein #18359.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 OS WO200175067-A2.
 PN WO200175067-A2.
 XX 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US008631.
 XX 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 DR P-PSDB; ABG18368.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

PS Claim 1; SEQ ID NO 18359; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64397-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2182 BP; 516 A; 569 C; 582 G; 515 T; 0 U; 0 Other;

Query Match 100.0%; Score 42; DB 5; Length 2182;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCACCTTACACATTCGTTAAGTCATATATGTTTGGACTTA 42
 |||||
 DB 487 CTGCACCTTACACATTCGTTAAGTCATATATGTTTGGACTTA 528

RESULT 4
 ACD28587/c
 ID ACD28587 standard; DNA; 40 BP.
 AC ACD28587;
 XX
 DT 10-OCT-2003 (first entry)
 XX
 DE E. coli Arsr binding oligonucleotide CHROML1B.
 DE
 KW Arsr; arsenic resistance operon; biosensor; ss; arsenic.
 XX
 OS Escherichia coli.
 XX
 PN US2003096275-A1.
 XX
 PD 22-MAY-2003.
 XX
 PF 15-AUG-2002; 2002US-00222952.
 XX
 PR 20-AUG-2001; 2001US-0313714P.
 XX
 PA (LAIN/) LAING L G.
 XX
 PI Laing LG;
 XX
 DR WPI; 2003-576876/54.
 XX

XX New system comprising isolated protein and nucleic acid, and a detection
 PT system that indicates a change in binding of the protein to the nucleic
 PT acid in the presence of the analyte, useful for detecting the presence of
 PT analyte in a sample.
 XX
 PS Claim 35; Page 15; 36pp; English.
 XX
 CC The invention relates to a new system (biosensor) for detecting the
 CC presence of analyte in a sample comprising: (1) an isolated protein that
 CC specifically binds the analyte; (2) an isolated nucleic acid containing a
 CC specific binding sequence that is bound specifically by the protein; and
 CC (3) a detection system that indicates a change in binding of the protein
 CC to the nucleic acid in the presence of the analyte. Also included are a
 CC biosensor device for detecting the presence of an analyte in a sample,
 CC detecting the presence of an analyte in a sample, and an Arsr (encoded by
 CC part of the arsenic resistance operon of E. coli) protein comprising an
 CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
 CC the Arsr sequence appearing as ABU63440 binding to a nucleic acid

XX
 SQ Sequence 25 BP; 13 A; 3 C; 2 G; 7 T; 0 U; 0 Other;
 Query Match 59.5%; Score 25; DB 8; Length 25;
 Best Local Similarity 100.0%; Pred. No. 10;

CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the bottom strand of an oligonucleotide which binds
 CC to a chromosomally expressed Arsr protein and is used in the biosensor of
 CC the invention

SQ Sequence 40 BP; 17 A; 5 C; 7 G; 11 T; 0 U; 0 Other;

Query Match 95.2%; Score 40; DB 8; Length 40;
 Best Local Similarity 100.0%; Pred. No. 4.2e-05;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GCACCTTACACATTCGTTAAGTCATATATGTTTGGACTTA 42
 |||||
 DB 40 GCACCTTACACATTCGTTAAGTCATATATGTTTGGACTTA 1

RESULT 5
 ACD28591/c
 ID ACD28591 standard; DNA; 25 BP.
 XX
 AC ACD28591;
 XX
 DT 10-OCT-2003 (first entry)
 XX
 DE E. coli Arsr binding oligonucleotide CHROMS1B.
 DE
 KW Arsr; arsenic resistance operon; biosensor; ss; arsenic.
 XX
 OS Escherichia coli.
 XX
 PN US2003096275-A1.
 XX
 PD 22-MAY-2003.
 XX
 PF 15-AUG-2002; 2002US-00222952.
 XX
 PR 20-AUG-2001; 2001US-0313714P.
 XX
 PA (LAIN/) LAING L G.
 XX
 PI Laing LG;
 XX
 DR WPI; 2003-576876/54.
 XX

XX New system comprising isolated protein and nucleic acid, and a detection
 PT system that indicates a change in binding of the protein to the nucleic
 PT acid in the presence of the analyte, useful for detecting the presence of
 PT analyte in a sample.
 XX
 PS Claim 35; Page 15; 36pp; English.
 XX
 CC The invention relates to a new system (biosensor) for detecting the
 CC presence of analyte in a sample comprising: (1) an isolated protein that
 CC specifically binds the analyte; (2) an isolated nucleic acid containing a
 CC specific binding sequence that is bound specifically by the protein; and
 CC (3) a detection system that indicates a change in binding of the protein
 CC to the nucleic acid in the presence of the analyte. Also included are a
 CC biosensor device for detecting the presence of an analyte in a sample,
 CC detecting the presence of an analyte in a sample, and an Arsr (encoded by
 CC part of the arsenic resistance operon of E. coli) protein comprising an
 CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
 CC the Arsr sequence appearing as ABU63440 binding to a nucleic acid

SQ Sequence 25 BP; 13 A; 3 C; 2 G; 7 T; 0 U; 0 Other;

Query Match 59.5%; Score 25; DB 8; Length 25;
 Best Local Similarity 100.0%; Pred. No. 10;

[illegible]

Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 6 CTTACACATTCCTTAAGTCATATATGTTTTCGACTTA 42
|||||
Db 6433 CTTACCAATAGTTATGGCATGTATGTTTATATGTA 6469
|||||

RESULT 13
ABLO5632/c
ID ABL05632 standard; cDNA; 3109 BP.
XX
AC ABL05632;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 11378.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW Pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
FN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEXE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB61529.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT Genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 11378; 21pp + Sequence Listing: English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC sequences (ABLO1840-ABLI6175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO.int/pub/published_pct_sequences
XX
SQ Sequence 3109 BP; 841 A; 699 C; 709 G; 860 T; 0 U; 0 Other;

Query Match 53.3%; Score 22.4; DB 4; Length 3109;
Best Local Similarity 72.5%; Pred. No. 1.4e+02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 TGCACCTTACACATTCGTTAAGTCATATATGTTTTCGACTT 41
|||||
Db 2682 TGTATATACATAGTCAATTTATTTATGTTTATTCCTT 2643
|||||

RESULT 14
ABQ71197/c
ID ABQ71197 standard; DNA; 1773 BP.
XX
AC ABQ71197;
XX
DT 29-AUG-2003 (revised)

DT 29-AUG-2002 (first entry)
XX
DE Listeria monocytogenes 4b specific contig124.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX
OS Listeria monocytogenes ATCC 19115.
XX
FN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR003061.
XX
PR 04-OCT-2000; 2000FR-00012697.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
DR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators.
XX
PS Claim 22; SEQ ID NO 4010; 180pp; French.
XX
CC The present invention relates to nucleic acid sequences (ABO67188-
CC ABO71212) from Listeria sp. The sequences are useful as probes and
CC primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of gene
CC expression. Proteins encoded by the nucleic acid sequences can be used to
CC screen for compounds that modulate gene expression, replication and
CC pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in anti-
CC Listeria vaccines. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO.int/pub/published_pct_sequences. (Updated
CC on 29-AUG-2003 to standardise OS field)
XX
SQ Sequence 1773 BP; 624 A; 223 C; 340 G; 537 T; 0 U; 49 Other;

Query Match 52.9%; Score 22.2; DB 6; Length 1773;
Best Local Similarity 77.1%; Pred. No. 1.6e+02;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 TGCACCTTACACATTCGTTAAGTCATATATGTTTTCGACTT 36
|||||
Db 1230 TGCCTTACCACATTCGTTTAGTCGCGATGTTTTCG 1196
|||||

RESULT 15
ABQ71020
ID ABQ71020 standard; DNA; 8148 BP.
XX
AC ABQ71020;
XX
DT 29-AUG-2003 (revised)
DT 29-AUG-2002 (first entry)
XX
DE Listeria monocytogenes 4b contig DNA sequence #962.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX
OS Listeria monocytogenes ATCC 19115.
XX
FN WO200228891-A2.
XX
PD 11-APR-2002.

XX 04-OCT-2001; 2001WO-FR003061.
 PF XX
 PR 04-OCT-2000; 2000FR-00012597.
 XX
 FA (INSP) INST PASTEUR.
 FA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Kurst F, Glaser P;
 XX
 DR WPI; 2002-332479/37.
 XX
 PT New genomic sequences from Listeria species, useful for detection,
 PT treatment and prevention of infection, also related polypeptides,
 PT antibodies and modulators.
 XX
 PS Claim 14; SEQ ID NO 3833; 180pp; French.
 XX
 CC The present invention relates to nucleic acid sequences (ABQ67188-
 CC ABQ71212) from Listeria sp. The sequences are useful as probes and
 CC primers for identification and/or detection of Listeria (e.g. as
 CC contaminants in foods, or mutational analysis) and for analysis of gene
 CC expression. Proteins encoded by the nucleic acid sequences can be used to
 CC screen for compounds that modulate gene expression, replication and
 CC pathogenicity of Listeria (potential therapeutic agents), also for
 CC treating infections by Listeria, and are useful as immunogens in anti-
 CC Listeria vaccines. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. (Updated
 CC on 29-AUG-2003 to standardise OS field)
 XX
 SQ Sequence 8148 BP; 2408 A; 1730 C; 1229 G; 2781 T; 0 U; 0 Other;
 Query Match 52.9%; Score 22.2; DB 6; Length 8148;
 Best Local Similarity 77.1%; Pred. No. 1.9e+02;
 Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 Oy 2 TGCACATTCGTTTACGTATATATGTTTT 16
 |||||
 Db 492 TGCCTACCATTCGTTTACGTATATGTTTT 526
 |||||

Search completed: May 26, 2004, 17:50:05
 Job time : 303.923 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 16:18:00 ; Search time 66.9487 Seconds
(without alignments)
348.146 Million cell updates/sec

Title: US-10-676-299-5

Perfect score: 42

Sequence: 1 cgcacttacacattggtta.....tcatatatgttttgactta 42

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/prodata/2/ina/5A COMB.seq.*
2: /cgn2_6/prodata/2/ina/5B COMB.seq.*
3: /cgn2_6/prodata/2/ina/6A COMB.seq.*
4: /cgn2_6/prodata/2/ina/6B COMB.seq.*
5: /cgn2_6/prodata/2/ina/PCUTS COMB.seq.*
6: /cgn2_6/prodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	50.0	192	4	US-09-134-001C-1096
C 2	21	50.0	711	4	US-09-601-198-111
C 3	21	50.0	1275	4	US-09-543-681A-1066
C 4	20.8	49.5	1317	4	US-09-499-203-15
C 5	20.8	49.5	2199	4	US-09-708-725A-3
C 6	20.8	49.5	4066	4	US-09-499-203-16
C 7	20.8	49.5	6204	4	US-09-499-203-20
C 8	20.8	49.5	7387	4	US-09-499-203-17
C 9	20.8	49.5	9321	4	US-09-499-203-1
C 10	20.6	49.0	8566	4	US-10-027-983-10
C 11	20.6	49.0	168375	4	US-09-426-290-1
C 12	20.4	48.6	342	2	US-08-607-384A-4
C 13	20.4	48.6	825	4	US-09-328-352-1636
C 14	20.4	48.6	2728	4	US-09-620-312D-572
C 15	20.4	48.6	5301	4	US-08-956-171E-443
C 16	20.4	48.6	35524	3	US-08-923-137-1
C 17	20.4	48.6	1830121	4	US-09-557-884-1
C 18	20.4	48.6	1830121	4	US-09-643-980A-1
C 19	20.2	48.1	270	4	US-09-543-681A-3761
C 20	20.2	48.1	651	4	US-09-134-001C-1749
C 21	20.2	48.1	689	4	US-09-495-050A-20
C 22	20.2	48.1	1419	4	US-09-540-236-177
C 23	20.2	48.1	3470	4	US-09-620-312D-571
C 24	20.2	48.1	10614	1	US-08-135-511-35
C 25	20.2	48.1	10614	1	US-08-187-453-35
C 26	20.2	48.1	58909	4	US-09-596-003-30
C 27	20.2	48.1	1664976	4	US-08-916-421B-1

ALIGNMENTS

RESULT 1

US-09-134-001C-1096
; Sequence 1096, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1096
; LENGTH: 192
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1096

Query Match 50.0%; Score 21; DB 4; Length 192;

Best Local Similarity 73.0%; Pred. NO. 46;

Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 TGCACCTACACATTCGTTAACTCATATATGTTTGA 38
DB 71 TGTATTCACATTAATCGTGAATCATTTATGACTTTGA 107

RESULT 2

US-09-601-198-111/c
; Sequence 111, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.
; APPLICANT: Chen, Ellison Y.
; APPLICANT: Glass, Jennifer S.
; APPLICANT: Glass, John I.
; APPLICANT: Heiner, Cheryl R.
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
; TITLE OF INVENTION: UREALYTICUM
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601,198
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181

Sequence 79, Appl
Sequence 22, Appl
Sequence 11, Appl
Sequence 112, Appl
Sequence 2373, Ap
Sequence 603, App
Sequence 7058, Ap
Sequence 1327, Ap
Sequence 1327, Ap
Sequence 1327, Ap
Sequence 16107, A
Sequence 3331, Ap
Sequence 198, App
Sequence 1, Appl
Sequence 1, Appl
Sequence 8, Appl


```
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 6204
; TYPE: DNA
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-20

Query Match      49.5%; Score 20.8; DB 4; Length 6204;
Best Local Similarity 70.0%; Pred. No. 78;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy      2  TGCACCTTACACATTCGTTAAGTCATATATGTTTTGACTT 41
Db      5663  TGACCATTCAAATTCGTCGAAGTAATATGTTATGCCGACTT 5624

RESULT 8
US-09-499-203-17/c
; Sequence 17, Application US/09499203
; Patent No. 6570065
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 7387
; TYPE: DNA
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-17

Query Match      49.5%; Score 20.8; DB 4; Length 7387;
Best Local Similarity 70.0%; Pred. No. 80;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy      2  TGCACCTTACACATTCGTTAAGTCATATATGTTTTGACTT 41
Db      6334  TGACCATTCAAATTCGTCGAAGTAATATGTTATGCCGACTT 6295

RESULT 9
US-09-499-203-1/c
; Sequence 1, Application US/09499203
; Patent No. 6570065
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 9321
; TYPE: DNA
; ORGANISM: Leuconostoc mesenteroides
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (678)..(6848)
US-09-499-203-1

Query Match      49.5%; Score 20.8; DB 4; Length 9321;
Best Local Similarity 70.0%; Pred. No. 82;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy      2  TGCACCTTACACATTCGTTAAGTCATATATGTTTTGACTT 41
Db      6334  TGACCATTCAAATTCGTCGAAGTAATATGTTATGCCGACTT 6295

RESULT 10
US-10-027-983-10/c
; Sequence 10, Application US/10027983
; Patent No. 6617162
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Mark P. Roach
; TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
; FILE REFERENCE: RTS-0340
; CURRENT APPLICATION NUMBER: US/10/027,983
; CURRENT FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 10
; LENGTH: 8566
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(2302)
; NAME/KEY: exon:exon junction
; LOCATION: (1491)...(1492)
; OTHER INFORMATION: exon 4:exon 5
; NAME/KEY: start codon
; LOCATION: (2302)...(2305)
; NAME/KEY: 3'UTR
; LOCATION: (2735)...(8566)
; NAME/KEY: exon:exon junction
; LOCATION: (3589)...(3570)
; OTHER INFORMATION: exon 8:exon 9
; NAME/KEY: exon:exon junction
; LOCATION: (3708)...(3709)
; OTHER INFORMATION: exon 9:exon 10
; NAME/KEY: exon:exon junction
; LOCATION: (3842)...(3843)
; OTHER INFORMATION: exon 10:exon 11
; NAME/KEY: exon:exon junction
; LOCATION: (4026)...(4027)
; OTHER INFORMATION: exon 11:exon 12
US-10-027-983-10

Query Match      49.0%; Score 20.6; DB 4; Length 8566;
Best Local Similarity 74.3%; Pred. No. 96;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy      1  CTGCACCTTACACATTCGTTAAGTCATATATGTTTTT 35
Db      3144  CTGCACACATATAGTCGTTATGTCATATATTTTTT 3110

RESULT 11
US-09-426-290-1/c
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berglind Ran Olafsdottir
; APPLICANT: Jeffrey Gulcher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
```

; STRAIN: ATCC 27958
 ; US-08-607-384A-4
 Query Match 48.6%; Score 20.4; DB 2; Length 342;
 Best Local Similarity 80.0%; Pred. No. 80;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 9 ACACATTCGTTAAGTCATATATGTTTTGA 38
 DB 14 ACACGTTGGTTAAGTCATATATGTTTTGA 43

RESULT 13
 US-09-328-352-1636/c
 ; Sequence 1636, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breston et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; CURRENT APPLICATION NUMBER: US/09/328.352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 1636
 ; LENGTH: 825
 ; TYPE: DNA
 ; ORGANISM: Acinetobacter baumannii
 ; US-09-328-352-1636

Query Match 48.6%; Score 20.4; DB 4; Length 825;
 Best Local Similarity 71.1%; Pred. No. 88;
 Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 5 ACTTACATTCGTTAAGTCATATATGTTTTGACTTA 42
 DB 546 ACTTCGAGAACCGATCAGTAAATTTGTTTCACGTA 509

RESULT 14
 US-09-620-312D-572/c
 ; Sequence 572, Application US/09620312D
 ; Patent No. 6569662
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Y. Tom
 ; APPLICANT: Liu, Chenghua
 ; APPLICANT: Asundi, Vinod
 ; APPLICANT: Zhang, Jie
 ; APPLICANT: Ren, Feiyan
 ; APPLICANT: Chen, Rui-hong
 ; APPLICANT: Zhao, Qing A.
 ; APPLICANT: Wehrman, Tom
 ; APPLICANT: Xue, Aidong J.
 ; APPLICANT: Yang, Yonghong
 ; APPLICANT: Wang, Jian-Rui
 ; APPLICANT: Zhou, Ping
 ; APPLICANT: Ma, Yungling
 ; APPLICANT: Wang, Dunrui
 ; APPLICANT: Wang, Zhiwei
 ; APPLICANT: John Tillinghast
 ; APPLICANT: Drmanac, Radjoe T.
 ; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
 ; FILE REFERENCE: Polypeptides
 ; FILE REFERENCE: 784CIF2B
 ; CURRENT APPLICATION NUMBER: US/09/620.312D
 ; CURRENT FILING DATE: 2000-07-19
 ; PRIOR APPLICATION NUMBER: 09/552,317
 ; PRIOR FILING DATE: 2000-04-25
 ; PRIOR APPLICATION NUMBER: 09/488,725
 ; PRIOR FILING DATE: 2000-01-21
 ; NUMBER OF SEQ ID NOS: 1105
 ; SOFTWARE: pt FL genes Version 1.0
 ; SEQ ID NO 572

```

; LENGTH: 2728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (202)..(1212)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2728)
; OTHER INFORMATION: n = a,t,c or g
US-09-620-312D-572

Query Match      48.6%; Score 20.4; DB 4; Length 2728;
Best Local Similarity 71.1%; Pred. No. 1e+02;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 5 ACTTACACATCGTTAAGTCATATATGTTTGGACTTA 42
    |||||
Db 1930 ACTGAACATAGTAGTATATATTTTAGATATA 1893

RESULT 15
US-08-956-171E-443
; Sequence 443, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 443:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5301 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 443:
US-08-956-171E-443

Query Match      48.6%; Score 20.4; DB 4; Length 5301;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;

```

```

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 11 ACATTGCTTAAGTCATATATGTTTGGACT 40
    |||||
Db 2324 ACATCCCTTAAGGCATATAATTTTGTCT 2353

```

Search completed: May 26, 2004, 17:56:40
Job time : 70.9487 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 16:21:09 ; Search time 2910.03 Seconds
(without alignments)
430.997 Million cell updates/sec

Title: US-10-676-299-5

Perfect score: 42

Sequence: 1 ctgcattacacattcgta.....tcataatgttttgactta 42

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mem:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	61.9	502	12	BM539131
2	24.8	59.0	233	12	BM307747
3	24.8	59.0	481	10	AW458707
4	24.8	59.0	525	12	BM568294

5	24.8	59.0	640	10	BB824180
6	24.4	58.1	379	29	CE496382
7	24.4	58.1	533	29	CE195732
8	24.4	58.1	621	29	CE212602
9	24.2	57.6	880	28	BZ388776
10	24	57.1	1064	28	BH174046
11	23.6	56.2	255	29	CG231937
12	23.6	56.2	420	28	AZ405411
13	23.6	56.2	480	29	CG197578
14	23.6	56.2	649	28	BH051848
15	23.6	56.2	703	28	BZ042702
16	23.6	56.2	905	28	CC400118
17	23.6	56.2	969	29	CG182590
18	23.6	56.2	1020	29	CG182586
19	23.4	55.7	211	14	CD525003
20	23.4	55.7	233	14	CD526210
21	23.4	55.7	286	13	BB529046
22	23.4	55.7	386	13	BY608251
23	23.4	55.7	448	9	AV518537
24	23.4	55.7	861	29	BX173966
25	23.4	55.7	874	29	CG971624
26	23.2	55.2	821	13	BQ429190
27	23	54.8	375	28	AQ137091
28	23	54.8	389	29	CE565867
29	23	54.8	600	28	BZ178813
30	23	54.8	762	29	CE008178
31	23	54.8	882	14	CD082216
32	23	54.8	948	29	CG952813
33	23	54.8	972	29	CG966166
34	23	54.8	1038	29	CG052815
35	22.8	54.3	322	10	BH120824
36	22.8	54.3	375	10	AW556632
37	22.8	54.3	386	14	W98216
38	22.8	54.3	467	10	BF320226
39	22.8	54.3	519	9	AA500057
40	22.8	54.3	543	14	CD292926
41	22.8	54.3	579	10	BH032011
42	22.8	54.3	588	28	BH015678
43	22.8	54.3	617	28	BH423034
44	22.8	54.3	639	28	BH315621
45	22.8	54.3	647	28	CC086257

ALIGNMENTS

RESULT 1
BM539131
LOCUS
DEFINITION
hh05c02.g1 Canis cDNAs from testes cells Canis familiaris CDNA
clone hh05c02 5', mRNA sequence.
BM539131
VERSION
KEYWORDS
SOURCE
ORGANISM
Canis familiaris (dog)
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE
1 (bases 1 to 502)
O'Shaughnessy, A.L., McCombie, W.R., Baker, J.P., Balija, V.,
Cunha, D., Dedhia, N., de la Bastide, M., Katzenberger, P.,
King, L., Kirchhoff, K.A., Miller, B., Muller, S., Nascimento, L.U.,
Palmer, L., Santos, L., Shah, R.S., Spiegel, L.A., Zutavern, T.,
Preston, R. and Hannon, G.J.
Expressed sequence tags from Canis familiaris (dog) (2002)
Unpublished (2002)
Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mcombie@cshl.org

/clone lib="Gm-cl075"
/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from mRNA isolated
from differentiating somatic embryos cultured on M6M6AC.
The library was prepared using the Stratagene pluescript
II SK(+) library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with an XhoI restriction site. EcoRI
adaptors were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA fragments were
directionally cloned into the EcoRI-XhoI restriction site
of the pBluescript vector. The ligated cDNA fragments
were transformed into E.coli Electromax DH10B host cells.
Tissue culture and library construction were performed by
Françoise Thibaud-Nissen and Anu Khanna (Lila Vodkin lab,
University of Illinois)."

	Query Match	59.0%;	Score 24.8;	DB 12;	Length 233;
	Best Local Similarity	80.6%;	Pred. No. 2.4e+02;		
	Matches 29;	Conservative 0;	Mismatches 7;	Indels 0;	Gaps 0;
OY	7	TTACACATTCGGTAAAGTCATATATCTTTTTCACCTTA	42		
Db	71	TAAATTCATTCGGTAAATCATATATATTTTCGACCTTA	36		

RESULT 3					
AW458707					
LOCUS	AW458707	481 bp	mRNA	linear	EST 03-DEC-2001
DEFINITION	sh12f03.y1 Gm-cl016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:				
	Gm-cl016-4566 5', mRNA sequence.				
ACCESSION	AW458707				

AW458707.1	GI:7028924
EST.	
KEYWORDS	Glycine max (soybean)
SOURCE	Glycine max
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosoids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE	1 (bases 1 to 481)

AUTHORS
Shoemaker, R., Keim, P., Vodkin, L., Erpelding, J., Corvett, V.,
Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,
Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,
Bowers, Y., Pearson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N.,
Schurr, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
McCann, R., Waterston, R., and Wilson, R.

TITLE
Public Soybean EST Project

JOURNAL
Unpublished (1993)

COMMENT
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Resgen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccu@resgen.com
Insert Length: 1046 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 411.
Location/Qualifiers

```

1. .481
source
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl016-4566"
/tissue_type="immature flowers of field grown plants"
/lab_host="XL10-Gold"
/clone_lib="Gm-cl016"
/note="Vector: pBluescript II KS; Site 1: EcoRI; Site 2:

```

XhoI; This cDNA library was constructed from mRNA isolated from immature flowers of field grown plants. The cDNA library was prepared using the Strategene pluscript II XR library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting."

ORIGIN

Query Match 59.0%; Score 24.8; DB 10; Length 481;
 Best Local Similarity 80.6%; Pred. No. 2.3e+02;
 Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 7 TTACACATTCGTTAAGTCATATATATGTTTTCGACTTA 42
 |||||
 Db 15 TAATTCATTCGCTAAATCATATATATTTTCGACTTA 50
 |||||

RESULT 4
 BM568294/c
 LOCUS
 DEFINITION BM568294 525 bp mRNA linear EST 21-FEB-2002
 Gm-c1057-4063 5', mRNA sequence.

ACCESSION BM568294
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Glycine max (soybean)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.

REFERENCE 1 (bases 1 to 525)
 AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V.,
 Khanna,A., Solla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
 Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
 Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
 Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
 McCann,R., Waterston,R. and Wilson,R.

Public Soybean EST Project
 Unpublished (1999)
 CONTACT: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: est@watson.wustl.edu
 This clone is available through: ResGen, Invitrogen Corp. 2130
 South Memorial Parkway Huntsville, AL 35801 For further information
 call: (800)-533-4363 or contact: ccu@resgen.com web site:
 www.resgen.com

Seq primer: -40RP from Gibco
 High quality sequence stop: 434.
 Location/Qualifiers

FEATURES

1..525
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="SOYBEAN CLONE ID: Gm-c1057-4063"
 /issue_type="Degenerating cotyledons, 2 week old seedling"
 /lab_host="DH108"
 /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2: XhoI. The cDNA library was constructed from mRNA isolated from degenerating cotyledons of 2 week old seedlings from

PI468916. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH108 host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

ORIGIN

Query Match 59.0%; Score 24.8; DB 12; Length 525;
 Best Local Similarity 80.6%; Pred. No. 2.3e+02;
 Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 7 TTACACATTCGTTAAGTCATATATATGTTTTCGACTTA 42
 |||||
 Db 423 TAATTCATTCGCTAAATCATATATATTTTCGACTTA 388
 |||||

RESULT 5
 BM56824180
 LOCUS
 DEFINITION BM56824180 640 bp mRNA linear EST 24-MAY-2001
 Gm-r1070 Glycine max cDNA clone Gm-r1070-8886 3', mRNA sequence.

ACCESSION BM56824180
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Glycine max (soybean)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.

REFERENCE 1 (bases 1 to 640)
 AUTHORS Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
 Erpelting,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
 A Functional Genomics Program for Soybean (NSF 9872565)
 Unpublished (1999)

Other ESTs: AW568532 corresponding to Gm-r1030-3256 (5')
 Contact: Vodkin, L.O., PI, A Functional Genomics Program for
 Soybean (NSF 9872565)

Lewin, H. A., Director, Keck Center for Comparative and Functional
 Genomics
 University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
 Tel: (217) 244-6147
 Fax: (217) 333-4582
 Email: l-vodkin@uiuc.edu

This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134. For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or info@genome
 systems.com web site: www.genomesystems.com

Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.
 Location/Qualifiers

FEATURES

1..640
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="Gm-r1070-8886"
 /note="The library Gm-r1070 is a sequence-driven, reracked set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons or a representative of each contig, which were reracked to form library Gm-r1070. The cDNA clones of the reracked Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by

the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, http://www.cbcb.umd.edu/ResearchProjects/Soybean/index.html . Reracking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois.

http://www.life.uiuc.edu/biotech/keck.html. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

ORIGIN

Query Match 59.0%; Score 24.8; DB 10; Length 640;
Best Local Similarity 80.6%; Pred. No. 2.3e+02;
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 7 TTACACATTCGTTAAAGTCATATATGTTTTCGACTTA 42
128 TAATTCATTCGCTAAATCATATATATTTTCGACTTA 163

RESULT 6
CE496382/c
LOCUS
DEFINITION tigr-gss-dog-1700037306768 Dog Library Canis familiaris genomic, genomic survey sequence.
ACCESSION CE496382
VERSION CE496382.1 GI:36813163
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
REFERENCE 1 (bases 1 to 379)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis. Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K., Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and Venter, J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
MEDLINE 22875432
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
Location/Qualifiers
1. .379
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from peripheral blood"

FEATURES
source

Query Match 59.1%; Score 24.4; DB 29; Length 379;
Best Local Similarity 73.8%; Pred. No. 3.1e+02;
Matches 31; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 CTGCACCTTACACATTCGTTAAAGTCATATATGTTTTCGACTTA 42
251 CTGCACATCCACATAAATGGAATCATATATGTTTTCGACTTA 210

ORIGIN

Query Match 58.1%; Score 24.4; DB 29; Length 379;
Best Local Similarity 73.8%; Pred. No. 3.1e+02;
Matches 31; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 CTGCACCTTACACATTCGTTAAAGTCATATATGTTTTCGACTTA 42
251 CTGCACATCCACATAAATGGAATCATATATGTTTTCGACTTA 210

RESULT 7

CE195732/c
LOCUS
DEFINITION tigr-gss-dog-17000371501174 Dog Library Canis familiaris genomic, genomic survey sequence.
ACCESSION CE195732
VERSION CE195732.1 GI:35351385
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
REFERENCE 1 (bases 1 to 533)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis. Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K., Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and Venter, J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
MEDLINE 22875432
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
Location/Qualifiers
1. .533
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from peripheral blood"

FEATURES
source

Query Match 58.1%; Score 24.4; DB 29; Length 533;
Best Local Similarity 73.8%; Pred. No. 3.1e+02;
Matches 31; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 CTGCACCTTACACATTCGTTAAAGTCATATATGTTTTCGACTTA 42
405 CTGCACATCCACATAAATGGAATCATATATGTTTTCGACTTA 364

ORIGIN

Query Match 58.1%; Score 24.4; DB 29; Length 533;
Best Local Similarity 73.8%; Pred. No. 3.1e+02;
Matches 31; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 CTGCACCTTACACATTCGTTAAAGTCATATATGTTTTCGACTTA 42
405 CTGCACATCCACATAAATGGAATCATATATGTTTTCGACTTA 364

RESULT 8

CE212602/c
LOCUS
DEFINITION tigr-gss-dog-1700037263569 Dog Library Canis familiaris genomic, genomic survey sequence.
ACCESSION CE212602
VERSION CE212602.1 GI:35368271
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
REFERENCE 1 (bases 1 to 621)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis. Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K., Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and Venter, J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
MEDLINE 22875432
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-0200

Fax: 301-838-0208
Email: ekirnes@igr.org
Class: shotgun.

FEATURES
source
Location/Qualifiers

1..621
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BatXI; Libraries were prepared from peripheral blood"

ORIGIN

Query Match 58.1%; Score 24.4; DB 29; Length 621;
Best Local Similarity 73.8%; Pred. No. 3.1e+02;
Matches 31; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 CTGCACATTCGTTAAGTCATATATGTTTTCACCTTA 42
|||||
Db 496 CTGCACATCCATAAATGATCATATATGTTTTCATTA 455
|||||

RESULT 9
BZ388776
LOCUS
DEFINITION BZ388776 880 bp DNA linear GSS 30-APR-2003
ENTAMEOBA INVADENS genomic clone E1012
GENOMIC SURVEY SEQUENCE.

ACCESSION BZ388776
VERSION BZ388776.1 GI:30235313
KEYWORDS GSS.

SOURCE Entamoeba invadens

ORGANISM Entamoeba invadens

REFERENCE Eukaryota; Entamoebidae; Entamoeba.

AUTHORS Wang, Z.; Samuelson, J., Clark, C.G., Eichinger, D., Paul, J., van

Deellen, K., Hall, N., Anderson, I., and Loftus, B.

TITLE Gene discovery in the Entamoeba invadens genome

JOURNAL Mol. Biochem. Parasitol. 129 (1), 23-31 (2003)

MEDLINE 22684048

PUBMED 12798503

COMMENT Other GSSs: E1012

CONTACT: Brendan Loftus

DEPARTMENT OF Eukaryotic Genomics

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-3543

Fax: 301-838-0208

Email: enta@igr.org

DNA was provided by Daniel Eichinger

Seq primer: 1F

Class: sheared ends.

Location/Qualifiers

1..880

/organism="Entamoeba invadens"

/mol_type="genomic DNA"

/strain="IP-1"

/db_xref="taxon:33085"

/clone="E1012"

/clone_lib="E1012"

/note="Vector: pPOS2; Site 1: BstXI; Total genomic DNA was isolated from early log phase trophozoites of E. invadens IP-1 using a Qiagen plant DNA extraction kit. A shotgun medium-size plasmid library (average insert size of 10-12 kb) was generated by random mechanical shearing of E. invadens genomic DNA, repairing the ends of DNA fragments with T4 Polymerase, adding BstXI adaptors and ligating into the BstXI site of a pUC-derived vector pPOS2."

ORIGIN

Query Match 57.6%; Score 24.2; DB 29; Length 880;
Best Local Similarity 78.4%; Pred. No. 3.5e+02;
Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 6 CTTCACATTCGTTAAGTCATATATGTTTTCACCTTA 42
|||||
Db 340 CTTCACAAATCGTTAAGACGGGTTGTTTTCACATA 376
|||||

RESULT 10
BH174046/c

LOCUS BH174046

DEFINITION

TOUS RFLP clone from soybean genomic DNA

GENOMIC SURVEY SEQUENCE.

ACCESSION BH174046

VERSION BH174046.1

KEYWORDS GSS.

SOURCE Glycine max (soybean)

ORGANISM Glycine max

REFERENCE Rukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

Glycine.

1 (bases 1 to 1064)

Grant, D. and Shoemaker, R.C.

unpublished sequence of soybean RFLP probe

Unpublished (2001)

CONTACT: Grant D

AGRONOMY DEPARTMENT

USDA-ARS and Iowa State University

G304 Agronomy Hall, Ames, IA 50011-1010, USA

Tel: 515 294 1205

Fax: 515 294 2299

Email: dgrant@iastate.edu

single pass sequence

Class: RFLP clone.

Location/Qualifiers

1..1064

/organism="Glycine max"

/mol_type="genomic DNA"

/db_xref="taxon:3847"

/clone="pf005"

/clone_lib="RFLP"

/note="Vector: pBS+; PstI-generated fragments of genomic

DNA. Library construction described by Keim, P. &

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."

COMMENT

Other GSSs: CGVDL62TH
 Contact: Cathy Whitelaw
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: TP
 Class: sheared ends.

FEATURES

source
 1..255
 Location/Qualifiers

/organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZM0511K04"
 /clone_lib="ZM 0.7-1.5 KB"
 /notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
 methylation filtered genomic DNA library"

ORIGIN

Query Match 56.2%; Score 23.6; DB 29; Length 255;
 Best Local Similarity 76.3%; Pred. No. 5.9e+02;
 Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 5 ACTTACACATTCGTTAGTCATATATATGTTTTCACCTA 42
 |||||
 Db 153 AATTATATATTCATTAAGTAATATATGATTTAATTTA 116

RESULT 12

AZ405411 420 bp DNA linear GSS 03-OCT-2000
 LOCUS
 DEFINITION
 clone UUGC1M0174A13 F, Genomic survey sequence.
 ACCESSION
 AZ405411
 VERSION
 AZ405411.1 GI:10529424
 KEYWORDS
 GSS.
 SOURCE
 Mus musculus (house mouse)

ORGANISM

Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 420)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,
 Reilly, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von
 Niederhausern, A. and Wright, D. Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

Plasmid inserts

Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177

Email: dunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0174 row: A column: 13
 Seq primer: CGTTGTAAACGACGCCACT
 Class: plasmid ends
 High quality sequence stop: 420.

FEATURES

source
 1..420
 Location/Qualifiers

/organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0174A13"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"

ORIGIN

Query Match 56.2%; Score 23.6; DB 29; Length 480;
 Best Local Similarity 76.3%; Pred. No. 5.7e+02;
 Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

/note="Vector: PWD42mv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PWD42 (GI:4732114|9b|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

ORIGIN

Query Match 56.2%; Score 23.6; DB 28; Length 420;
 Best Local Similarity 76.3%; Pred. No. 5.8e+02;
 Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 CACTTACACATTCGTTAGTCATATATGTTTTCACCTT 41
 |||||
 Db 256 CACTTACACATTCGTTAGTCATATATGTTTTCACCTT 293

RESULT 13

CG197578 480 bp DNA linear GSS 21-AUG-2003
 LOCUS
 DEFINITION
 PUFMC47TD ZM_0.6-1.0 KB Zea mays genomic clone ZM0681H21,
 Genomic survey sequence.

ACCESSION
 CG197578
 VERSION
 CG197578.1 GI:34088639
 KEYWORDS
 GSS.

ORGANISM

Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

1 (bases 1 to 480)
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
 Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
 Bennetzen, J.

TITLE

Maize Genomics Consortium

JOURNAL

Unpublished (2003)

COMMENT

Other GSSs: PUFMC47TB
 Contact: Cathy Whitelaw
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: TP
 Class: sheared ends.

FEATURES

source
 1..480
 Location/Qualifiers

/organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZM0681H21"
 /clone_lib="ZM 0.6-1.0 KB"
 /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
 Cot selected genomic DNA library"

```

5  ACTTACACATTCGTTAAGTCATATATGTTTTGACTTA 42
81  AATCACACATTCATTAAAGTAATATATGAATGAATTA 118

RESULT 14
BH051848/c
LOCUS
DEFINITION
    BH051848          649 bp      DNA          linear      GSS 17-JUL-2001
    RPCI-24-26007.TJ RPCI-24 Mus musculus genomic clone RPCI-24-26007,
    genomic survey sequence.
ACCESSION
    BH051848
VERSION
    BH051848.1  GI:14943837
KEYWORDS
    GSS.
SOURCE
    Mus musculus (house mouse)
ORGANISM
    Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
    1  (bases 1 to 649)
    Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akınret,B., Levins,M.,
    Thegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,B.,
    Russell,D., de Jong,P. and Fraser,C.M.
    Mouse BAC End Sequences from Library RPCI-24
    Unpublished (1999)
    Other_GSSs: RPCI-24-26007.TV
CONTACT: Shaying Zhao
    Department of Eukaryotic Genomics
    The Institute for Genomic Research
    9712 Medical Center Dr., Rockville, MD 20850, USA
    Tel: 301 838 0200
    Fax: 301 838 0208
    Email: szhaoc@tigr.org
    Clones are derived from the mouse BAC library RPCI-24. For BAC
    library availability, please contact Pieter de Jong
    (pdejong@email.cho.org). Clones may be purchased from BACPAC
    resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
    pages: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
    Plate: 260 row: O column: 7
    Seq primer: SP6
    Class: BAC ends.
    Location/Qualifiers
        1. .649
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
FEATURES
    source

```

```

ORIGIN
Query Match          56.2%;   Score 23.6;   DB 28;   Length 649;
Best Local Similarity 76.3%;   Pred. No. 5.7e+03;
Matches 29;   Conservative 0;   Mismatches 9;   Indels 0;   Gaps 0;

QY      5  ACTTACACATTGGTAAAGTCATATATGTTTTGACTTA 42
          ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      466 AGTTCAGTTTCCTTCAGTCATTATGTTGAGACTTA 429

RESULT 15
BZ042702/c
LOCUS      BZ042702
DEFINITION 1jq75c05.b1 S.oleracea002 Brassica oleracea genomic, genomic survey
sequence.
ACCESSION  BZ042702
VERSION    BZ042702.1  GI:23634332

DNA. "

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 15:22:28 ; Search time 285.641 Seconds
(without alignments)
594.900 Million cell updates/sec

Title: US-10-676-299-6

Perfect score: 40

Sequence: 1 taagtcacaaacatatgacttaacgaatgtgaagtgc 40

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 337863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747736

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 29Jan04:*

1: geneseqn1980s:*

2: geneseqn1990s:*

3: geneseqn2000s:*

4: geneseqn2001as:*

5: geneseqn2001bs:*

6: geneseqn2002as:*

7: geneseqn2002bs:*

8: geneseqn2003bs:*

9: geneseqn2003cs:*

10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	40	100.0	40	8	ACD28587	Ac28587 E. coli A
2	40	100.0	42	8	ACD28586	Ac28586 E. coli A
3	40	100.0	401	7	ACA15469	ACA15469 Prokaryot
4	40	100.0	2182	5	AA828555	AA828555 DNA encod
5	25	62.5	25	8	ACD28591	Ac28591 E. coli A
6	25	62.5	25	8	ACD28590	Ac28590 E. coli A
7	22.8	57.0	396	6	ABN25854	ABn25854 Human ORF
8	22.6	56.5	6872	4	ABU18112	Ab118112 Drosophil
9	22	55.0	3547	4	ABU11816	Ab111816 Drosophil
10	21.8	54.5	3741	4	ABU05790	Ab105790 Drosophil
11	21.8	54.5	3856	4	ABU05792	Ab105792 Drosophil
12	21.8	54.5	19082	6	ABU32626	Ab132626 Human imm
13	21.6	54.0	3503	4	ABU21302	Ab121302 Drosophil
14	21.6	54.0	15667	6	ABU34146	Ab134146 Human imm
15	21.6	54.0	15933	4	ABU30408	Ab130408 Drosophil
16	21.6	54.0	17929	9	ADC86000	Adc86000 Human GPC
17	21.6	54.0	41104	6	AA36260	Aad36260 Human G-p
18	21.4	53.5	534	5	ABV51050	Abv51050 Human pro
19	21.4	53.5	660	6	ABQ21634	Abq21634 Oligonuc1
20	21.4	53.5	660	6	ABQ21635	Abq21635 Oligonuc1
21	21.4	53.5	1719	6	AB232449	Ab232449 Candida a
22	21.4	53.5	1733	3	AAC45407	Aac45407 Arabidops
23	21.4	53.5	1848	3	AAC45408	Aac45408 Arabidops

24	21.4	53.5	3109	4	ABL05632	Ab105632 Drosophil
c 25	21.4	53.5	29956	8	ADA02963	Ada02963 Mouse Lck
c 26	21.4	53.5	29956	9	ADB72701	Adb72701 Mouse Lck
c 27	21.4	53.5	29956	9	ADC85443	Adc85443 Mouse Lck
c 28	21.2	53.0	1287	7	ACF70667	Acf70667 Phototrab
29	21.2	53.0	1773	6	ABQ71197	Abq71197 Listeria
30	21.2	53.0	7823	4	AAS45490	Aas45490 Chemical
31	21.2	53.0	7823	6	ABL34061	Ab134061 Human imm
32	21.2	53.0	7823	6	ABK31493	Abk31493 Signal tr
33	21.2	53.0	7823	6	ABK28418	Abk28418 DNA trans
c 34	21.2	53.0	8148	6	ABQ71020	Abq71020 Listeria
35	21.2	53.0	110000	7	ACF67367	Acf67367_37
c 36	21.2	53.0	110000	7	ACF65388	Acf65388_10
c 37	21	52.5	347	7	ABX51820	Abx51820 Bovine ES
38	21	52.5	711	2	AAX99611	Aax99611 Nucleic a
39	21	52.5	2535	7	ACA47576	AcA47576 Prokaryot
c 40	21	52.5	5082	4	ABL04626	Ab104626 Drosophil
c 41	21	52.5	54786	8	ADA02705	Ada02705 Mouse Zfh
c 42	21	52.5	54786	9	ADB72443	Adb72443 Mouse Zfh
43	21	52.5	110000	6	ABX08336	Abx08336_04
44	21	52.5	177563	8	ACD28257	Ac28257 Mouse sol
45	21	52.5	325791	4	AAS43104	Aas43104 Human Oes

ALIGNMENTS

RESULT 1
ACD28587
ID ACD28587 standard; DNA; 40 BP.
XX
AC ACD28587;
XX
DT 10-OCT-2003 (first entry)
XX
DE E. coli ArsrR binding oligonucleotide CHROML1B.
XX
KW ArsrR; arsenic resistance operon; biosensor; ss; arsenic.
XX
OS Escherichia coli.
XX
FN US2003096275-A1.
XX
PD 22-MAY-2003.
XX
PF 15-AUG-2002; 2002US-002222952.
XX
PR 20-AUG-2001; 2001US-0313714P.
XX
PA (LAIN/) LAING L G.
XX
PI Laing LG;
XX
DR WPI; 2003-576876/54.
XX
PT New system comprising isolated protein and nucleic acid, and a detection system that indicates a change in binding of the protein to the nucleic acid in the presence of the analyte, useful for detecting the presence of analyte in a sample.
XX
PS Claim 35; Page 15; 36pp; English.
XX
CC The invention relates to a new system (biosensor) for detecting the presence of analyte in a sample comprising: (1) an isolated protein that specifically binds the analyte; (2) an isolated nucleic acid containing a specific binding sequence that is bound specifically by the protein; and (3) a detection system that indicates a change in binding of the protein to the nucleic acid in the presence of the analyte. Also included are a biosensor device for detecting the presence of an analyte in a sample, detecting the presence of an analyte in a sample, and an ArsrR (encoded by part of the arsenic resistance operon of E. coli) protein comprising an amino acid sequence that is at least 90% identical to amino acids 1-97 of the ArsrR sequence appearing as ABU63440 binding to a nucleic acid

CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
CC system is useful for detecting the presence of analyte in a sample. The
CC present sequence is the bottom strand of an oligonucleotide which binds
CC to a chromosomally expressed Arsr protein and is used in the biosensor of
CC the invention
XX
SQ Sequence 40 BP; 17 A; 5 C; 7 G; 11 T; 0 U; 0 Other;

Query Match 100.0%; Score 40; DB 8; Length 40;
Best Local Similarity 100.0%; Pred. No. 4.9e-05;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TAAGTCAAAACATATATGACTTACGATGTTAAAGTGC 40
Db 1 TAAGTCAAAACATATATGACTTACGATGTTAAAGTGC 40

RESULT 2
ACD28586/c
ID ACD28586 standard; DNA; 42 BP.
XX AC ACD28586;
XX
DT 10-OCT-2003 (first entry)
XX
DE E. coli Arsr binding oligonucleotide CHROMLIT.
XX
KW Arsr; arsenic resistance operon; biosensor; ss; arsenic.
XX
OS Escherichia coli.
XX
PN US2003096275-A1.
XX
XX 22-MAY-2003.
XX
XX 15-AUG-2002; 2002US-00222952.
XX
XX 20-AUG-2001; 2001US-0313714P.
XX
XX (LAIN/) LAING L G.
XX
XX Laing LG;
XX
XX WPI; 2003-576876/54.
XX
XX New system comprising isolated protein and nucleic acid, and a detection
PT system that indicates a change in binding of the protein to the nucleic
PT acid in the presence of the analyte, useful for detecting the presence of
PT analyte in a sample.
XX
XX Claim 35; Page 15; 36pp; English.

CC The invention relates to a new system (biosensor) for detecting the
CC presence of analyte in a sample comprising: (1) an isolated protein that
CC specifically binds the analyte; (2) an isolated nucleic acid containing a
CC specific binding sequence that is bound specifically by the protein; and
CC (3) a detection system that indicates a change in binding of the protein
CC to the nucleic acid in the presence of the analyte. Also included are a
CC biosensor device for detecting the presence of an analyte in a sample,
CC detecting the presence of an analyte in a sample, and an Arsr (encoded by
CC part of the arsenic resistance operon of E. coli) protein comprising an
CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
CC the Arsr sequence appearing as AB063440 binding to a nucleic acid
CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
CC system is useful for detecting the presence of analyte in a sample. The
CC present sequence is the top strand of an oligonucleotide which binds to a
CC chromosomally expressed Arsr protein and is used in the biosensor of the
CC invention
XX
SQ Sequence 42 BP; 11 A; 8 C; 5 G; 18 T; 0 U; 0 Other;

Query Match 100.0%; Score 40; DB 8; Length 42;
Best Local Similarity 100.0%; Pred. No. 4.9e-05;

Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TAAGTCAAAACATATATGACTTACGATGTTAAAGTGC 40
Db 42 TAAGTCAAAACATATATGACTTACGATGTTAAAGTGC 3

RESULT 3
ACA15469
ID ACA15469 standard; DNA; 401 BP.
XX AC ACA15469;
XX
DT 27-OCT-2003 (revised)
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene antisense oligonucleotide #3339.
XX
KW Antisense; ss; prokaryotic essential gene; cell proliferation;
KW drug design.
XX
XX Archaea.
XX
XX WO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX 06-SEP-2001; 2001US-00948993.
XX
XX 25-OCT-2001; 2001US-0342923P.
XX
XX 08-FEB-2002; 2002US-00072851.
XX
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wail D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.

XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 1; SEQ ID NO 3339; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the 6213
CC antisense sequences of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 401 BP; 113 A; 58 C; 101 G; 119 T; 0 U; 0 Other;

Query Match 100.0%; Score 40; DB 7; Length 401;
Best Local Similarity 100.0%; Pred. No. 5.7e-05;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAGTCAAAACATATATGACTTAACGAATGTGTAAAGTGC 40
|||
DB 104 TAAGTCAAAACATATATGACTTAACGAATGTGTAAAGTGC 143

RESULT 4
AAS82555/c
ID AAS82555 standard; cDNA; 2182 BP.

XX AAS82555;
AC
XX
DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #18359.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSB-) HYSB INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX P-PSDB; ABG18368.

XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 1; SEQ ID NO 18359; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have application in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and

CC amino acid sequences, AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 2182 BP; 516 A; 569 C; 582 G; 515 T; 0 U; 0 Other;
Query Match 100.0%; Score 40; DB 5; Length 2182;
Best Local Similarity 100.0%; Pred. No. 6.4e-05;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAGTCAAAACATATATGACTTAACGAATGTGTAAAGTGC 40
|||
DB 528 TAAGTCAAAACATATATGACTTAACGAATGTGTAAAGTGC 489

RESULT 5
ACD28591
ID ACD28591 standard; DNA; 25 BP.

XX ACD28591;
AC
XX
DT 10-OCT-2003 (first entry)
DE *E. coli* ArsR binding oligonucleotide CHROMS1B.
XX
XX ArsR; arsenic resistance operon; biosensor; ss; arsenic.

XX Escherichia coli.
XX
XX US2003096275-A1.
XX
XX 22-MAY-2003.
XX
XX 15-AUG-2002; 2002US-00222952.
XX
XX 20-AUG-2001; 2001US-0313714P.
XX
XX (LAIN/) LAING L G.

XX Laing LG;
XX
XX WPI; 2003-576876/54.

XX New system comprising isolated protein and nucleic acid, and a detection
XX system that indicates a change in binding of the protein to the nucleic
XX acid in the presence of the analyte, useful for detecting the presence of
XX analyte in a sample.

XX Claim 35; Page 15; 36pp; English.

XX The invention relates to a new system (biosensor) for detecting the
XX presence of analyte in a sample comprising: (1) an isolated protein that
XX specifically binds the analyte; (2) an isolated nucleic acid containing a
XX specific binding sequence that is bound specifically by the protein; and
XX (3) a detection system that indicates a change in binding of the protein
XX to the nucleic acid in the presence of the analyte. Also included are a
XX biosensor device for detecting the presence of an analyte in a sample,
XX detecting the presence of an analyte in a sample, and an ArsR (encoded by
XX part of the arsenic resistance operon of *E. coli*) protein comprising an
XX amino acid sequence that is at least 90% identical to amino acids 1-97 of
XX the ArsR sequence appearing as ABU63440 binding to a nucleic acid
XX sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
XX system is useful for detecting the presence of analyte in a sample. The
XX present sequence is the bottom strand of an oligonucleotide which binds
XX to a chromosomally expressed ArsR protein and is used in the biosensor of
XX the invention

XX Sequence 25 BP; 13 A; 3 C; 2 G; 7 T; 0 U; 0 Other;

Query Match 62.5%; Score 25; DB 8; Length 25;
Best Local Similarity 100.0%; Pred. No. 11;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAGTCAAAAACATATGACTTAA 25
|||||
Db 1 TAAGTCAAAAACATATGACTTAA 25

RESULT 6
ACD28590/C
ID ACD28590 standard; DNA; 25 BP.
XX
AC ACD28590;
XX
DT 10-OCT-2003 (first entry)
XX
DE E. coli ArsR binding oligonucleotide CHROMSIT.
XX
KW ArsR; arsenic resistance operon; biosensor; ss; arsenic.
XX
OS Escherichia coli.
XX
FN US2003096275-A1.
XX
PD 22-MAY-2003.
XX
PF 15-AUG-2002; 2002US-00222952.
XX
PR 20-AUG-2001; 2001US-0313714P.
XX
PA (LAIN/) LAING L G.
XX
PI Laing LG;
XX
DR WPI; 2003-576876/54.
XX

New system comprising isolated protein and nucleic acid, and a detection PT
PT system that indicates a change in binding of the protein to the nucleic PT
PT acid in the presence of the analyte, useful for detecting the presence of PT
PT analyte in a sample.
XX
PS Claim 35; Page 15; 36pp; English.
XX
CC The invention relates to a new system (biosensor) for detecting the CC
CC presence of analyte in a sample comprising: (1) an isolated protein that CC
CC specifically binds the analyte; (2) an isolated nucleic acid containing a CC
CC specific binding sequence that is bound specifically by the protein; and CC
CC (3) a detection system that indicates a change in binding of the protein CC
CC to the nucleic acid in the presence of the analyte. Also included are a CC
CC biosensor device for detecting the presence of an analyte in a sample, CC
CC detecting the presence of an analyte in a sample, and an ArsR (encoded by CC
CC part of the arsenic resistance operon of E. coli) protein comprising an CC
CC amino acid sequence that is at least 90% identical to amino acids 1-97 of CC
CC the ArsR sequence appearing as AB063440 binding to a nucleic acid CC
CC sequence comprising oligonucleotides appearing as ACD28594-ACD28591. The CC
CC system is useful for detecting the presence of analyte in a sample. The CC
CC present sequence is the top strand of an oligonucleotide which binds to a CC
CC chromosomally expressed ArsR protein and is used in the biosensor of the CC
CC invention
XX
SQ Sequence 25 BP; 7 A; 2 C; 3 G; 13 T; 0 U; 0 Other;

Query Match 62.5%; Score 25; DB 8; Length 25;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAGTCAAAAACATATGACTTAA 25
|||||
Db 25 TAAGTCAAAAACATATGACTTAA 1

RESULT 7
ABN25854
ID ABN25854 standard; cDNA; 396 BP.

XX
AC ABN25854;
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX polynucleotide sequence SEQ ID NO:20185.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis; gene; ss.
XX
OS Homo sapiens.
XX
FN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US010836.
XX
PR 30-MAY-2000; 2000US-0206132P.
XX
PR 29-AUG-2000; 2000US-0228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;
XX
DR WPI; 2002-106308/14.
XX
DR P-PSDB; ABP10102.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders.
XX
PS Disclosure; SEQ ID NO 20185; 1037pp; English.

XX
CC The present invention describes substantially purified human proteins CC
CC (referred to as open reading frame ORFX, where X is 1-11491 (see Table 1 CC
CC in the specification). ABN15762 to ABN27232 encode the human ORFX CC
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for CC
CC treating or preventing a pathology associated with an ORFX-associated CC
CC disorder in humans, and in the manufacture of a medicament for treating a CC
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide CC
CC sequences can be used in gene therapy. ORFX sequences can be used in the CC
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver, CC
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, CC
CC osteoarthritis, neurodegenerative disorders, disorders related to organ CC
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic CC
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester CC
CC storage disease, various immune deficiencies and disorders, infectious CC
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid CC
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host CC
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also CC
CC useful for treating burns, incisions, ulcers, for treating osteoporosis, CC
CC bone degenerative disorders, or periodontal disease, and for gut CC
CC reperfusion or regeneration and treatment of lung or liver fibrosis, CC
CC systemic cytokine damage. N.B. The sequence data for this patent did not CC
CC form part of the printed specification, but was obtained in electronic CC
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 396 BP; 148 A; 81 C; 51 G; 116 T; 0 U; 0 Other;

Query Match 57.0%; Score 22.8; DB 6; Length 396;
Best Local Similarity 79.4%; Pred. No. 82;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 AGTCAAAAACATATGACTTACGATGTGTA 36
|||||

Dd 272 AGTTAATAACATTTTGTCTTTACAAATGTTAA 305

RESULT 8
ABL18112/c
ID ABL18112 standard; DNA; 6872 BP.
XX
XX ABL18112;
AC
XX
XX Df
Df 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 5809.
DE
DE Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
KW
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US009231.
PF
XX 23-MAR-2000; 2000US-0191637P.
PR
XX 11-JUL-2000; 2000US-00614150.
PR
XX (PEKE) PE CORP NY.
PA
XX Venter JC, Adams M, Li PWD, Myers EM;
PI WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 5809; zipp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 6872 BP; 2081 A; 1490 C; 1437 G; 1864 T; 0 U; 0 Other;

Query Match 56.5%; Score 22.6; DB 4; Length 6872;
Best Local Similarity 75.7%; Pred. No. 1.2e+02;
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0

Oy 1 TAAGTCAAAACATATATGCATCTTAACGAATGTGTAAG 37
||| ||||||| ||| ||| |||||||
Db 6469 TACATAATAAACATACATGCCATAACTATTGTGTAAG 6433

RESULT 9
ABL11816
ID ABL11816 standard; cDNA; 3647 BP.
XX
XX ABL11816;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 29930.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
KW

PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR P-PSDB; ABB61687.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 11852; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3741 BP; 951 A; 843 C; 856 G; 1091 T; 0 U; 0 Other;
XX
Query Match 54.5%; Score 21.8; DB 4; Length 3741;
Best Local Similarity 78.8%; Pred. No. 2.2e+02;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
OY 6 CAAAACATATATGACTTAACGATGTGTAAAGT 38
DB 680 CATTAACCTTAATGACTTAAGGATTGTGTAAT 648
RESULT 11
ABL05792/C
ID ABL05792 standard; cDNA; 3856 BP.
XX
AC ABL05792;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 11858.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW Pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
FN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB61689.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 11858; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and

CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3856 BP; 973 A; 906 C; 882 G; 1095 T; 0 U; 0 Other;
XX
Query Match 54.5%; Score 21.8; DB 4; Length 3856;
Best Local Similarity 78.8%; Pred. No. 2.2e+02;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
OY 6 CAAAACATATATGACTTAACGATGTGTAAAGT 38
DB 3550 CATTAACCTTAATGACTTAAGGATTGTGTAAT 3518
RESULT 12
ABL32626/C
ID ABL32626 standard; DNA; 19082 BP.
XX
AC ABL32626;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 599.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianemic; cytostatic; neotropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
FN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 599; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 19082 BP; 5097 A; 429 C; 4336 G; 9220 T; 0 U; 0 Other;
XX
Query Match 54.5%; Score 21.8; DB 6; Length 19082;

Best Local Similarity 78.8%; Pred. No. 2.5e+02;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 6 CAAAAACATATATGACTTAACGAATGTGAAGT 38
DB 7388 CAAAAAATATAAACTTAACAAATTTATATAAT 7356

RESULT 13
ABL21302
ID: ABL21302 standard; DNA; 3503 BP.

XX AC ABL21302;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 15379.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX PT interactions.
XX PS Claim 1; SEQ ID NO 15379; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

QY 5 TCAGAAACATATATGACTTAACGAATGTGAAGTGC 40
DB 912 TCAGAAACATATATGACTTAACGAATGTGAAGTGC 947

RESULT 14
ABL34146/c
ID: ABL34146 standard; DNA; 15667 BP.
XX AC ABL34146;
XX DT 26-MAR-2002 (first entry)

XX DE Human immune system associated gene SEQ ID NO: 2119.
XX KW Human; immune system disease; cytosine methylation; antiasthmatic;
XX KW antiarteriosclerotic; antianaemic; cytosstatic; nootropic;
XX KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX KW ds.
XX OS Homo sapiens.
XX PN WO200200928-A2.
XX PD 03-JAN-2002.
XX PF 02-JUL-2001; 2001WO-EP007537.
XX PR 30-JUN-2000; 2000DE-01032529.
XX PR 01-SEP-2000; 2000DE-01043826.
XX PA (EPIC-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-130909/17.
XX PT Nucleic acid comprising fragment of chemically modified gene, useful for
XX PT diagnosis and treatment of diseases associated with abnormal cytosine
XX PT methylation.
XX PS Claim 1; SEQ ID NO 2119; 32pp + Sequence Listing; German.
XX CC The present invention provides a number of human immune system associated
XX CC genes which are modified by the methylation of cytosines. The sequences
XX CC can be used in the diagnosis and treatment of immune system disorders,
XX CC including eye diseases such as retinopathy, neovascular glaucoma and
XX CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX CC diseases. The present sequence is a gene of the invention
XX SQ Sequence 15667 BP; 4544 A; 186 C; 3395 G; 7552 T; 0 U; 0 Other;
Query Match 54.0%; Score 21.6; DB 6; Length 15667;
Best Local Similarity 75.0%; Pred. No. 2.9e+02;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 TAAGTCAAAACATATATGACTTAACGAATGTGAAT 36
DB 3417 TAACCAAAAACCTAAATTTACTTAACAAATTTATTA 3382

RESULT 15
ABL30408
ID: ABL30408 standard; DNA; 15933 BP.
XX AC ABL30408;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 42697.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.

```

XX 23-MAR-2001; 2001WO-US009231.
PF
XX 23-MAR-2000; 2000US-0191637P.
XX
PR 11-JUL-2000; 2000US-00614150.
PR
XX (PEXE ) PE CORP NY.
XX
PA Venter JC, Adams M, Li FWD, Myers EW;
XX
PI WPI; 2001-656860/75.
XX
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 42697; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
CC sequences (AB101840-ABL16175) and the encoded proteins (ABBS7737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 15933 BP; 4230 A; 3492 C; 3488 G; 4723 T; 0 U; 0 Other;
SQ
Query Match 54.0%; Score 21.6; DB 4; Length 15933;
Best Local Similarity 75.0%; Pred. No. 2.9e+02;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 TAAATCAGAAACATATATGACTTACGATGTGTA 36
Db 11838 TTAGGTAAACAATATCTTACTAAACGATATGTAA 11873
Search completed: May 26, 2004, 17:50:07
Job time : 287.641 secs

```

Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID S
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 1056

```
; LENGTH: 1275
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-1066

Query Match          52.5%; Score 21; DB 4; Length 1275;
Best Local Similarity 73.0%; Pred. No. 51;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy  2  AAGTCAAAACATATATGACTTAACGAATGTGTAAGT 38
    |||||  |||||  |||||  |||||  |||||  |||||
Db  678  AGGTAAAGAGATATAATAATTAAGAGAGGTGAAGT 714

RESULT 3
US-09-426-290-1
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berglind Ran Olafsdottir
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)
; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
US-09-426-290-1

Query Match          51.5%; Score 20.6; DB 4; Length 168575;
Best Local Similarity 74.3%; Pred. No. 1,1e+02;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy  1  TAAGTCAAAACATATATGACTTAACGAATGTGTA 35
    |||||  |||||  |||||  |||||  |||||  |||||
Db  118232  TATTGTCAATCATATATGTGTTAAGGAATTGTA 118266

RESULT 4
US-08-607-384A-4/c
; Sequence 4, Application US/08607384A
; Patent No. 5849488
; GENERAL INFORMATION:
; APPLICANT: ALATOSSAVA, JOUKO TAPANI
; APPLICANT: FORSMAN, P IVI TUULIKKI
; APPLICANT: TILSALA-TIMESJ RVI, ANU KYLLIKKI
; TITLE OF INVENTION: DNA SEQUENCE-BASED DIAGNOSIS OF MASTITIS
; TITLE OF INVENTION: FROM A MILK SAMPLE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
```

```
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/607,384A
; FILING DATE: 27-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 227-75
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus uberis
; STRAIN: ATCC 27958
US-08-607-384A--4

Query Match          51.0%; Score 20.4; DB 2; Length 342;
Best Local Similarity 80.0%; Pred. No. 75;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy  5  TCATAAACATATATGACTTAACGAATGTGT 34
    |||||  |||||  |||||  |||||  |||||  |||||
Db  43  TCAAACTAAATAAGACTTAACCAACGTGT 14

RESULT 5
US-09-328-352-1636
; Sequence 1636, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1636
; LENGTH: 825
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1636

Query Match          51.0%; Score 20.4; DB 4; Length 825;
Best Local Similarity 71.1%; Pred. No. 81;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy  1  TAAGTCAAAACATATATGACTTAACGAATGTGTAAGT 38
    |||||  |||||  |||||  |||||  |||||  |||||
Db  509  TACGTGAAAACAATTTTACTGATCGTTCGTAAGT 546

RESULT 6
US-09-708-725A-3
; Sequence 3, Application US/09708725A
; Patent No. 6489456
; GENERAL INFORMATION:
; APPLICANT: LADUNGA et al.
```



```

; APPLICANT: Fisher, Krishna J.
; TITLE OF INVENTION: Chimpanzee Adenovirus Vectors
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House and Corporate Cntr., P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: United States of America
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,137
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,700
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GKNPN.021CIP1USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35524 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; US-08-923-137-1

```

```

Query Match 51.0%; Score 20.4; DB 3; Length 35524;
Best Local Similarity 71.1%; Pred. No. 1.1e+02;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```

```

Qy 2 AAGTCAAAACATATATGACTTAACGAATGTGTAAGTG 39
|||
Db 20014 AAACAAACACCTATGACTACATGATGGCGGTG 20051
|||

```

```

RESULT 10
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8439
; TELEFAX: 301-309-8504
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1
Query Match 51.0%; Score 20.4; DB 4; Length 1830121;
Best Local Similarity 71.1%; Pred. No. 1.2e+02;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
Qy 1 TAAGTCAAAACATATATGACTTAACGAATGTGTAAGT 38
|||
Db 854524 TACGTCAAAATGAAGTATGATTAATGAATTTTAAAT 854561
|||
RESULT 11
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs

```

```
;
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match          51.0%; Score 20.4; DB 4; Length 1830121;
Best Local Similarity 71.1%; Pred. No. 1.2e+02;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 TAAGTCAAAAACATATATGACTTAACGAATGTGTAAAGT 38
   ||||||| ||||||| ||||||| ||||||| |||||||
DB 854524 TAGTCAAAATGAAGTATGATTAAATGAATTTTAAAT 854561

RESULT 12
US-09-543-681A-3761/c
; Sequence 3761, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 3344
; SEQ ID NO 3761
; LENGTH: 270
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-3761

Query Match          50.5%; Score 20.2; DB 4; Length 270;
Best Local Similarity 75.8%; Pred. No. 87;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TAAGTCAAAAACATATATGACTTAACGAATGTG 33
   ||||||| ||||||| ||||||| ||||||| |||||||
DB 176 TAAGTAAAAAATAAATGCTTTATAATGTG 144

RESULT 13
US-09-134-001C-1096/c
; Sequence 1096, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1096
; LENGTH: 192
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1096

Query Match          50.0%; Score 20; DB 4; Length 192;
Best Local Similarity 72.2%; Pred. No. 99;
Matches 26; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 5 TCAGTCAAAAACATATGCTTTAAGCAATGTGTAAAGTCC 40
   ||||||| ||||||| ||||||| ||||||| |||||||
DB 107 TCAAGTCAAAAATGAATGATTTCAGGATTATGTGAATC 72
```

```
RESULT 14
US-08-861-774E-79/c
; Sequence 79, Application US/08861774E
; Patent No. 6297007
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miso, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; FILE REFERENCE: 9953-006
; CURRENT APPLICATION NUMBER: US/08/861,774E
; CURRENT FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 79
; LENGTH: 761
; TYPE: DNA
; ORGANISM: Usnea florida
US-08-861-774E-79

Query Match          50.0%; Score 20; DB 3; Length 761;
Best Local Similarity 72.2%; Pred. No. 1.1e+02;
Matches 26; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTCAAAACATATATGACTTAACGAATGTGTAAAGT 39
   ||||||| ||||||| ||||||| ||||||| |||||||
DB 427 GCCTAAAACATTTATTATTATTAAGAGGATGACGTG 392

RESULT 15
US-09-499-203-15
; Sequence 15, Application US/09499203
; Patent No. 6570065
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 1917
; TYPE: DNA
; ORGANISM: leuconostoc mesenteroides
US-09-499-203-15

Query Match          50.0%; Score 20; DB 4; Length 1917;
Best Local Similarity 72.2%; Pred. No. 1.2e+02;
Matches 26; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 AAGTCAAAAACATATATGACTTAACGAATGTGTAAAG 37
   ||||||| ||||||| ||||||| ||||||| |||||||
DB 1433 AAGTCGGCATAACATATATTACTTACGAAATTTGAATG 1468

Search completed: May 26, 2004, 17:56:45
Job time : 68.7607 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 16:21:09 ; Search time 2771.45 Seconds
430.997 Million cell updates/sec

Title: US-10-676-299-6

Perfect score: 40

Sequence: 1 taagtcacaaacatatatgacttaacgaatgtgaagtgc 40

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estlin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hcc:*

9: gb_estl:*

10: gb_est2:*

11: gb_hcc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gssl:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24.8	62.0	233	12 BM307747	BM307747 sak33d01.
c 2	24.8	62.0	481	10 AW458707	AW458707 sh12f03.y
3	24.8	62.0	525	12 BM568294	BM568294 sal02e04.
c 4	24.8	62.0	640	10 BB824180	BB824180 GM700023A

C	5	24.2	60.5	890	28	BZ388776
C	6	24	60.0	502	12	BM539131
C	7	23.6	59.0	255	29	CG231937
C	8	23.6	59.0	420	28	AZ405411
C	9	23.6	59.0	480	29	CG197578
C	10	23.6	59.0	649	28	BH051848
C	11	23.6	59.0	703	28	BZ042702
C	12	23.6	59.0	905	28	CC400118
C	13	23.6	59.0	969	29	CG182590
C	14	23.6	59.0	1020	29	CG182586
C	15	23.4	58.5	1064	28	BH174046
C	16	23	57.5	375	28	AQ137091
C	17	23	57.5	389	29	CE565867
C	18	23	57.5	861	29	BX173966
C	19	23	57.5	948	29	CG052813
C	20	23	57.5	972	29	CG966166
C	21	23	57.5	1038	29	CG052815
C	22	22.8	57.0	579	10	BM032011
C	23	22.8	57.0	588	28	BH015678
C	24	22.8	57.0	1201	9	AL532945
C	25	22.6	56.5	202	28	BZ751549
C	26	22.6	56.5	202	28	BZ751554
C	27	22.6	56.5	240	13	BQ871477
C	28	22.6	56.5	313	9	AW017417
C	29	22.6	56.5	491	28	BH079313
C	30	22.6	56.5	529	29	CC944471
C	31	22.6	56.5	594	29	CE216547
C	32	22.6	56.5	676	28	BH878660
C	33	22.6	56.5	696	28	BH569150
C	34	22.6	56.5	697	28	BH955873
C	35	22.6	56.5	699	29	CC856781
C	36	22.6	56.5	700	29	CB423849
C	37	22.6	56.5	762	29	CE008178
C	38	22.6	56.5	764	29	CG153250
C	39	22.6	56.5	789	28	BH207086
C	40	22.6	56.5	804	29	CG438832
C	41	22.6	56.5	823	28	BH542358
C	42	22.6	56.5	836	28	CC066189
C	43	22.6	56.5	848	29	CG173594
C	44	22.6	56.5	861	29	CG153249
C	45	22.6	56.5	882	29	CC690601

ALIGNMENTS

RESULT 1
BM307747
LOCUS
DEFINITION sak33d01.y1 Gm-c1075 Glycine max cdna clone SOYBEAN CLONE ID: linear EST 02-JAN-2002
Gm-c1075-4777 5', mRNA sequence.
BM307747
BM307747.1 GI:18039453
VERSION EST.
KEYWORDS Glycine max (soybean)
SOURCE Glycine max
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
1 (bases 1 to 233)

REFERENCE

AUTHORS

Shoemaker, R., Keim, P., Vodkin, L., Erpelding, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE

Public Soybean EST Project

Unpublished (1999)

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: ResGen, Invitrogen Corp. 2130
 South Memorial Parkway Huntville, AL 35801 For further information
 call: (800)-533-4363 or contact: ccu@resgen.com web site:
 www.resgen.com
 Seq primer: -40RP from Gibco
 High quality sequence stop: 223.

FEATURES

1. 233
 Location/Qualifiers
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="SOYBEAN CLONE ID: Gm-c1075-4777"
 /tissue_type="differentiating somatic embryos cultured on
 MSM6AC"
 /lab_host="DH10B"
 /clone_lib="Gm-c1075"
 /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
 XhoI; The cDNA library was constructed from mRNA isolated
 from differentiating somatic embryos cultured on MSM6AC.
 The library was prepared using the Stratagene pBluescript
 II SK(+) library construction kit. Complementary DNA was
 synthesized from mRNA using a primer consisting of a
 poly(dT) sequence with an XhoI restriction site. EcoRI
 adapters were ligated to the blunt-ended cDNA fragments
 followed by XhoI digestion. The cDNA fragments were
 directionally cloned into the EcoRI-XhoI restriction site
 of the pBluescript vector. The ligated cDNA fragments
 were transformed into E.coli ElectroMax DH10B host cells.
 Tissue culture and library construction were performed by
 Francoise Thibaud-Nissen and Anu Khana (Uila Vodkin lab,
 University of Illinois)."

ORIGIN

Query Match 62.0%; Score 24.8; DB 12; Length 233;
 Best Local Similarity 80.6%; Pred. No. 2.1e+02;
 Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 TAAGTCAAAACATATATGACTTACGAAATGTGTA 36
 |||||
 Db 36 TAAGTCGAAATATATATGATTAGCGAATGAATTA 71
 |||||

RESULT 2
 AW458707/c
 LOCUS sh12f03.y1 Gm-c1016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 DEFINITION Gm-c1016-4566 5', mRNA sequence.
 ACCESSION AW458707
 VERSION AW458707.1 GI:7028924
 KEYWORDS EST.
 SOURCE Glycine max (soybean)
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.

REFERENCE 1 (bases 1 to 481)
 AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V.,
 Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
 Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
 Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
 Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
 McCann,R., Waterston,R. and Wilson,R.
 Public Soybean EST Project
 Unpublished (1999)

TITLE Contact: Shoemaker R/Public Soybean EST Project
 JOURNAL Public Soybean EST Project
 COMMENT Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800

Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: ResGen, Invitrogen Corp. 2130
 South Memorial Parkway Huntville, AL 35801 For further information
 call: (800)-533-4363 or contact via email: ccu@resgen.com
 Insert length: 1046 Std Error: 0.00
 Seq primer: -40RP from Gibco
 High quality sequence stop: 411.

FEATURES

1. 481
 Location/Qualifiers
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1016-4566"
 /tissue_type="Immature flowers of field grown plants"
 /lab_host="XL10-Gold"
 /clone_lib="Gm-c1016"
 /note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2:
 XhoI; This cDNA library was constructed from mRNA isolated
 from immature flowers of field grown plants. The cDNA
 library was prepared using the Stratagene pBluescript II
 XR library construction kit. Complementary DNA was
 synthesized from mRNA using a primer consisting of a poly
 (dT) sequence with a XhoI restriction site. EcoRI adapters
 were ligated to the blunt-ended cDNA fragments followed by
 XhoI digestion. The cDNA fragments were directionally
 cloned into the EcoRI-XhoI restriction site of the
 pBluescript vector. The ligated cDNA fragments were
 transformed into XL10-Gold host cells. This library was
 constructed by Dr. Randy Shoemaker and Dr. John
 Erpelting."

ORIGIN

Query Match 62.0%; Score 24.8; DB 10; Length 481;
 Best Local Similarity 80.8%; Pred. No. 2e+02; 7; Indels 0; Gaps 0;
 Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 TAAGTCAAAACATATATGACTTACGAAATGTGTA 36
 |||||
 Db 50 TAAGTCGAAATATATATGATTAGCGAATGAATTA 15
 |||||

RESULT 3
 BM568294
 LOCUS sal02e04.y1 Gm-c1057 Glycine max cDNA clone SOYBEAN CLONE ID:
 DEFINITION Gm-c1057-4063 5', mRNA sequence.

ACCESSION BM568294
 VERSION BM568294.1 GI:18849186
 KEYWORDS EST.
 SOURCE Glycine max (soybean)
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.

REFERENCE 1 (bases 1 to 525)
 AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V.,
 Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
 Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
 Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
 Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
 McCann,R., Waterston,R. and Wilson,R.
 Public Soybean EST Project
 Unpublished (1999)

TITLE Contact: Shoemaker R/Public Soybean EST Project
 JOURNAL Public Soybean EST Project
 COMMENT Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: ResGen, Invitrogen Corp. 2130

South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact: ccu@resgen.com web site:
www.resgen.com

Seq primer: -40RP from Gibco
High quality sequence stop: 434.
Location/Qualifiers

FEATURES

1..525
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl057-4063"
/tissue type="Degenerating cotyledons, 2 week old seedling"
/lab host="DH10B"
/clone lib="Gm-cl057"
/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from mRNA isolated from degenerating cotyledons of 2 week old seedlings from PI468916. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

ORIGIN

Query Match 62.0%; Score 24.8; DB 12; Length 525;
Best Local Similarity 80.6%; Pred. No. 2e+02; 7; Indels 0; Gaps 0;
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 TAAGTCAAAACATATATGACTTACGAATGTGTAA 36
|||||
DB 368 TAAGTCGAATATATATGATTACGAATGAATTA 423
|||||

RESULT 4
BB824180/c
LOCUS BB824180.1 GI:10256414
DEFINITION GM700023A20B3 Gm-r1070 Glycine max cDNA clone Gm-r1070-8886 3', mRNA sequence.
ACCESSION BB824180
VERSION BB824180.1 GI:10256414
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
1 (bases 1 to 640)
Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V., Expelling,J., Rapp,C., Shoop,C., Pardinas,J., Liu,L. and Lewin,H. A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished (1999)
Other_ESTs: AW568532 corresponding to Gm-r1030-3256 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomeystems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES

source
1..640
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="Gm-r1070-8886"
/clone lib="Gm-r1070"
/note="The library Gm-r1070 is a sequence-driven, reracked set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig, which were reracked to form library Gm-r1070. The cDNA clones of the reracked Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html . Reracking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois
http://www.life.uiuc.edu/biotech/keck.html. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

ORIGIN

Query Match 62.0%; Score 24.8; DB 10; Length 640;
Best Local Similarity 80.6%; Pred. No. 2e+02; 7; Indels 0; Gaps 0;
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 TAAGTCAAAACATATATGACTTACGAATGTGTAA 36
|||||
DB 163 TAAGTCGAATATATATGATTACGAATGAATTA 128
|||||

RESULT 5
BZ388776/c
LOCUS BZ388776
DEFINITION EINBR83TF EI_10_12_KB Entamoeba invadens genomic clone EINBR83, genomic survey sequence.
ACCESSION BZ388776
VERSION BZ388776.1 GI:30235313
KEYWORDS GSS.
SOURCE Entamoeba invadens
ORGANISM Entamoeba invadens
Eukaryote; Entamoebidae; Entamoeba.
1 (bases 1 to 880)
Wang,Z., Samuelson,J., Clark,C.G., Eichinger,D., Paul,J., van Dellen,K., Hall,N., Anderson,I. and Loftus,B. Gene discovery in the Entamoeba invadens genome
Mol. Biochem. Parasitol. 129 (1), 23-31 (2003)
22684348
12798503
Other_GSSs: EINBR83TR
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: entactigr.org
DNA was provided by Daniel Eichinger
Seq primer: TF
Class: sheared ends.
Location/Qualifiers

FEATURES

```

source
1. .880
/organism="Entamoeba invadens"
/mol_type="Genomic DNA"
/strain="IP-1"
/db_xref="taxon:33085"
/clone="EINBR83"
/clone_lib="E1_10.12_KB"
/notes="Vector: pHOS2; Site 1: BstXI; Total genomic DNA was isolated from early log phase trophozoites of E. invadens IP-1 using a Qiagen plant DNA extraction kit. A shotgun medium-size plasmid library (average insert size of 10-12 kb) was generated by random mechanical shearing of E. invadens genomic DNA, repairing the ends of DNA fragments with T4 Polymerase, adding BstXI adaptors and ligating into the BstXI site of a pUC-derived vector pHOS2."

ORIGIN
Query Match 60.5%; Score 24.2; DB 28; Length 880;
Best Local Similarity 78.4%; Pred. No. 3.1e+02;
Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 TAAGTCAAAACATATATGACTTAACGATGTGTAA 37
Db 376 TATGTAATAAACAAACCCCTCTTAACGATTTGTAAG 340

RESULT 6
BM539131/c
LOCUS BM539131
DEFINITION hb05c02.g1 Canis cDNAs from testes cells Canis familiaris cDNA
ACCESSION BM539131
VERSION
KEYWORDS EST.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
REFERENCE 1 (bases 1 to 502)
AUTHORS O'Shaughnessy,A.L., McCombie,W.R., Baker,J.P., Balija,V., Cummins,D., Dedhia,N.N., de la Bastide,M., Katzenberger,F., King,L., Kirchoff,K.A., Miller,B., Muller,S., Nascimento,L.U., Palmer,L., Santos,L., Shah,R.S., Spiegel,L.A., Zutavern,T., Preston,R. and Hannon,G.J.
TITLE Expressed sequence tags from Canis familiaris (dog) (2002)
JOURNAL Unpublished (2002)
COMMENT Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mcombie@cshl.org
Plate: hb05 row: c column: 02
Seq primer: -21M3UnivRev
High quality sequence stop: 502.
Location/Qualifiers
1. .502
/organism="Canis familiaris"
/mol_type="mRNA"
/db_xref="taxon:9615"
/clone="hb05c02"
/tissue_type="testes"
/clone_lib="Canis cDNAs from testes cells"
/notes="Vector: Lambda Zap II; The library was produced by Greg Hannon and Raymond Preston (Cold Spring Harbor Laboratory). This library is oligo(dT) primed using stragene zap cDNA synthesis kit. It was made from dog testes. Please contact Greg Hannon (hannon@cshl.org) with any library related inquiries."

ORIGIN
Query Match 60.0%; Score 24; DB 12; Length 502;

Best Local Similarity 75.0%; Pred. No. 3.8e+02;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 TAAGTCAAAACATATATGACTTAACGATGTGTAACTGC 40
Db 238 TAATTGATACAAATTTATGACTTTAAGGAATGTAGAAGTTC 199

RESULT 7
CG231937
LOCUS CG231937
DEFINITION CGVLD62TV.ZM.0.7.1.5.KB.Zea mays genomic clone ZMMBMA0511K04, genomic survey sequence.
ACCESSION CG231937
VERSION CG231937.1 GI:34131823
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
REFERENCE 1 (bases 1 to 255)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Other GSSs: OGVLD62TH
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: Tg
Class: Sheared ends.
Location/Qualifiers
1. .255
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0511K04"
/clone_lib="ZM.0.7.1.5.KB"
/notes="Vector: pBCSK; Site 1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"

ORIGIN
Query Match 59.0%; Score 23.6; DB 29; Length 255;
Best Local Similarity 76.3%; Pred. No. 5.3e+02;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 TAAGTCAAAACATATATGACTTAACGATGTGTAACT 38
Db 116 TAAATTAAATTCATATATTAATTAATGATATATATT 153

RESULT 8
AZ405411/c
LOCUS AZ405411
DEFINITION 1M0174A13P Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0174A13 F, genomic survey sequence.
ACCESSION AZ405411
VERSION AZ405411.1 GI:10529424
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 420)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 420)
Dunn,D., Aoyagi,A., Barber,M., Seacore,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von

```

Niederhauser, A. and Wright, D., Weis, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0174 row: A column: 13
 Seq primer: CGTTGTAAACGACGCGCCACT
 Class: plasmid ends
 High quality sequence stop: 420.
 Location/Qualifiers
 1..420
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0174A13"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (g14732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

ORIGIN
 Query Match 59.0%; Score 23.6; DB 28; Length 420;
 Best Local Similarity 76.3%; Pred. No. 5.1e+02;
 Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 2 AAGTCAAAACATATATGACTTAAACGAATGTGTAAGTG 39
 DB 293 AAGTAAAAATAACTTGAACCTCAGGAATGTGTAAGTG 256

RESULT 9
 CG197578/c
 LOCUS CG197578 480 bp DNA linear GSS 21-AUG-2003
 DEFINITION PUFMC47TD ZM 0.6.1.0 KB Zea mays genomic clone ZMMBT00601H21,
 genomic survey sequence.
 ACCESSION CG197578
 VERSION CG197578.1 GI:34088639
 KEYWORDS GSS
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 480)
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
 Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and

FEATURES
 source
 1..480
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone_lib="ZM 0.6.1.0 KB"
 /clone="ZMMBT00601H21"
 /note="Vector: PCR4-TOPO, Site_1: EcoRI; 0.6-1.0 kb high
 Cot selected genomic DNA library"

ORIGIN
 Query Match 59.0%; Score 23.6; DB 29; Length 480;
 Best Local Similarity 76.3%; Pred. No. 5.1e+02;
 Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 1 TAAGTCAAAACATATATGACTTAAACGAATGTGTAAGT 38
 DB 118 TAAATACATTCATATATTAATTAATTAATGAATGTGTAAT 81

RESULT 10
 BH051848
 LOCUS BH051848 649 bp DNA linear GSS 17-JUL-2001
 DEFINITION RPCI-24-26007.TJ RPCI-24 Mus musculus genomic clone RPCI-24-26007,
 genomic survey sequence.
 ACCESSION BH051848
 VERSION BH051848.1 GI:14843837
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 649)
 Zhao, S., Nierman, W., Malek, J., Shvartsbeyn, A., Gebregorgis, E.,
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Other GSSs: RPCI-24-26007.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
 page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
 Plate: 260 row: O column: 7
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers
 1..649
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"

FEATURES
 source
 1..649
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"

FEATURES
 source
 1..480
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone_lib="ZM 0.6.1.0 KB"
 /clone="ZMMBT00601H21"
 /note="Vector: PCR4-TOPO, Site_1: EcoRI; 0.6-1.0 kb high
 Cot selected genomic DNA library"

ORIGIN
 Query Match 59.0%; Score 23.6; DB 29; Length 480;
 Best Local Similarity 76.3%; Pred. No. 5.1e+02;
 Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 1 TAAGTCAAAACATATATGACTTAAACGAATGTGTAAGT 38
 DB 118 TAAATACATTCATATATTAATTAATTAATGAATGTGTAAT 81

FEATURES
 source
 1..480
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone_lib="ZM 0.6.1.0 KB"
 /clone="ZMMBT00601H21"
 /note="Vector: PCR4-TOPO, Site_1: EcoRI; 0.6-1.0 kb high
 Cot selected genomic DNA library"

ORIGIN
 Query Match 59.0%; Score 23.6; DB 29; Length 480;
 Best Local Similarity 76.3%; Pred. No. 5.1e+02;
 Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 1 TAAGTCAAAACATATATGACTTAAACGAATGTGTAAGT 38
 DB 118 TAAATACATTCATATATTAATTAATTAATGAATGTGTAAT 81

Bennetzen, J.
 Maize Genomics Consortium
 Unpublished (2003)
 Other GSSs: PUFMC47TB
 Contact: Cathy Whitelaw
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: TF
 Class: sheared ends.
 Location/Qualifiers
 1..480
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone_lib="ZM 0.6.1.0 KB"
 /clone="ZMMBT00601H21"
 /note="Vector: PCR4-TOPO, Site_1: EcoRI; 0.6-1.0 kb high
 Cot selected genomic DNA library"

FEATURES

source

1..480
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone_lib="ZM 0.6.1.0 KB"
 /clone="ZMMBT00601H21"
 /note="Vector: PCR4-TOPO, Site_1: EcoRI; 0.6-1.0 kb high
 Cot selected genomic DNA library"

ORIGIN

Query Match 59.0%; Score 23.6; DB 29; Length 480;
 Best Local Similarity 76.3%; Pred. No. 5.1e+02;
 Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 1 TAAGTCAAAACATATATGACTTAAACGAATGTGTAAGT 38
 DB 118 TAAATACATTCATATATTAATTAATTAATGAATGTGTAAT 81

RESULT 10

BH051848

LOCUS BH051848

DEFINITION

RPCI-24-26007.TJ RPCI-24 Mus musculus genomic clone RPCI-24-26007,

genomic survey sequence.

ACCESSION BH051848

VERSION BH051848.1 GI:14843837

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 649)

Zhao, S., Nierman, W., Malek, J., Shvartsbeyn, A., Gebregorgis, E.,

Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,

Russell, D., de Jong, P. and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-24

Unpublished (1999)

Other GSSs: RPCI-24-26007.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end

page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html

Plate: 260 row: O column: 7

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..649

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

```

/clone="RPC1-24-26007"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPC1-24"
/note="Vector: PTARBAC1; Site 1: BamH1; Site 2: BamH1;
RPC1-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the PTARBAC1 cloning vector at the
BamH1 sites using MboI partially digested male C57BL/6J
DNA."

```

ORIGIN

```

Query Match      59.0%; Score 23.6; DB 28; Length 649;
Best Local Similarity 76.3%; Pred. No. 5e+02;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 TAAGTCAAAACATATGACTTACGAAATGTGAAGT 38
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 429 TAAGTCAAAACATATGACTTACGAAATGTGAAGT 466

```

```

RESULT 11
BZ042702
LOCUS BZ042702 703 bp DNA linear GSS 09-OCT-2002
DEFINITION BZ042702.1 B.oleracea002 Brassica oleracea genomic, genomic survey
Sequence.
ACCESSION BZ042702
VERSION BZ042702.1 GI:23634332
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea

```

```

REFERENCE 1 (bases 1 to 703)
AUTHORS Delehaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,
Nash, W., Rabinowicz, P.D. and Wilson, R.K.
TITLE Whole genome shotgun reads from Brassica oleracea
JOURNAL Unpublished (2002)
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Plate: 1j76 row: c column: 05
Seq primer: -2luppot forward
Class: shotgun
High quality sequence start: 94
High quality sequence stop: 551.

```

FEATURES

```

source
1..703
Location/Qualifiers
    /organism="Brassica oleracea"
    /mol_type="genomic DNA"
    /db_xref="taxon:3712"
    /clone_lib="B.oleracea002"
    /note="Vector: pOTW13; Whole genome shotgun library from
    flowering buds. DNA was purified from a crude nuclear
    prep using Brassica oleracea T01000DH3 buds provided by
    Thomas Osborn at the University of Wisconsin. Genomic
    DNA was provided by Pablo Rabinowicz (CSHL) and the
    shotgun library prepared at Washington University Genome
    Sequencing Center."

```

ORIGIN

```

Query Match      59.0%; Score 23.6; DB 28; Length 703;
Best Local Similarity 76.3%; Pred. No. 5e+02;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 TAAGTCAAAACATATGACTTACGAAATGTGAAGT 38
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 570 TAAGTCAAAACATATGACTTACGAAATGTGAAGT 607

```

```

RESULT 12
CC400118/c

```

```

LOCUS CC400118
DEFINITION PUHPP62TB ZM_0.6_1.0_KB Zea mays genomic clone ZMMB7a505K04,
genomic survey sequence.
ACCESSION CC400118
VERSION CC400118.1 GI:30880208
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays

```

```

REFERENCE 1 (bases 1 to 905)
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

```

```

TITLE Zea mays
JOURNAL Zea mays
COMMENT Zea mays
Maize Genomics Consortium
Unpublished (2003)
Other_GSSs: PUHPP62TD
Contact: Cathy Whitelaw
TIGR

```

```

9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.

```

FEATURES

```

source
1..905
Location/Qualifiers
    /organism="Zea mays"
    /mol_type="genomic DNA"
    /strain="B73"
    /db_xref="taxon:4577"
    /clone="ZMMB7a505K04"
    /clone_lib="ZM_0.6_1.0_KB"
    /note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
    Cot selected genomic DNA library"

```

ORIGIN

```

Query Match      59.0%; Score 23.6; DB 28; Length 905;
Best Local Similarity 76.3%; Pred. No. 5e+02;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 1 TAAGTCAAAACATATGACTTACGAAATGTGAAGT 38
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 339 TAAGTCAATTCATATATTACTTAATGAATGTGATT 302

```

```

RESULT 13
CG182590/c

```

```

LOCUS CG182590
DEFINITION PUKCN17TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMB7a0788C10,
genomic survey sequence.
ACCESSION CG182590
VERSION CG182590.1 GI:34073651
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays

```

```

REFERENCE 1 (bases 1 to 969)
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

```

```

TITLE Zea mays
JOURNAL Zea mays
COMMENT Zea mays
Maize Genomics Consortium
Unpublished (2003)
Other_GSSs: PUKCN17TB
Contact: Cathy Whitelaw
TIGR

```

```

9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org

```

RESULT 15

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 16:21:09 ; Search time 1593.59 Seconds
(without alignments)
430.997 Million cell updates/sec

Title: US-10-676-299-7

Perfect score: 23

Sequence: 1 ttaatcatatgcgtttttggta 23

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: em_estba:*

2: em_esthum:*

3: em_estlin:*

4: em_estm:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hic:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.4	84.3	774	29	CC924663
2	19	82.6	775	29	BX134638
C 3	18.8	81.7	616	28	AZ753174
C 4	18.8	81.7	594	29	CG009365

5	18.8	81.7	752	28	AQ788301
6	18.8	81.7	844	29	CC814073
C 7	18.8	81.7	973	28	CC419273
8	18.8	81.7	1101	29	CNS0020H
C 9	18.4	80.0	354	28	AZ045495
C 10	18.4	80.0	475	28	AZ045618
C 11	18.4	80.0	491	28	AZ045411
C 12	18.4	80.0	574	28	AQ579547
C 13	18.4	80.0	655	28	AQ579574
C 14	18.2	79.1	309	12	BM164586
15	18.2	79.1	343	9	AI738836
16	18.2	79.1	357	28	BZ094929
17	18.2	79.1	359	13	BX111464
18	18.2	79.1	400	28	BH370025
C 19	18.2	79.1	500	12	BI878636
C 20	18.2	79.1	523	28	AQ035362
C 21	18.2	79.1	539	12	BM156934
C 22	18.2	79.1	544	9	AI666913
C 23	18.2	79.1	550	10	AW344027
24	18.2	79.1	571	9	AV382514
C 25	18.2	79.1	582	12	BM857902
C 26	18.2	79.1	584	28	AQ016199
C 27	18.2	79.1	587	12	EG307569
C 28	18.2	79.1	601	12	BM184130
C 29	18.2	79.1	602	12	BM095920
C 30	18.2	79.1	608	12	BM183874
31	18.2	79.1	659	28	BZ113063
32	18.2	79.1	668	28	BH514860
33	18.2	79.1	709	29	CG928923
34	18.2	79.1	727	12	B1926826
35	18.2	79.1	744	28	BZ052628
36	18.2	79.1	768	12	BM171079
C 37	18.2	79.1	807	28	BZ822499
38	18.2	79.1	814	28	BH376658
C 39	18.2	79.1	830	28	AZ897507
40	18.2	79.1	845	28	BZ822494
41	17.8	77.4	275	10	BB373536
42	17.8	77.4	280	10	BB420422
43	17.8	77.4	478	10	BF449847
C 44	17.8	77.4	498	9	AA575651
45	17.8	77.4	515	12	BM226209

ALIGNMENTS

RESULT 1
CC924663
LOCUS
DEFINITION t078e02ba.fl TAMET Bos taurus genomic clone t078e02ba, genomic survey sequence.
ACCSSION
VERSION CC924663.1 GI:33560002
KEYWORDS
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae; Bovinae; Bos.
1 (bases 1 to 774)
Lin,S., Najjar,F.Z., Adelson,D., Gill,C.A. and Roe,B.A.
Bovine BAC End Sequences from Library TAMET
Unpublished (2003)
Contact: Bruce A. Roe
Advanced Center for Genome Technology
University of Oklahoma Department of Chemistry and Biochemistry
620 Farrington Oval, Room 208, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
Class: BAC ends
High quality sequence start: 39
High quality sequence stop: 668.


```

FEATURES
  source
    Location/Qualifiers
      1..774
        /organism="Bos taurus"
        /mol_type="genomic DNA"
        /strain="Angus bull T A M U Shoshone Y6 11519666"
        /db_xref="taxon:9913"
        /clone="t078e02ba"
        /sex="Male"
        /cell_type="Blood"
        /clone_lib="TAMBT"
        /note="Vector: pBelOBAC11; Site 1: HindIII; Site 2: HindIII; TAMBT Bovine BAC library (Male) produced by Texas A&M University, Department of Animal Science."
ORIGIN
  Query Match      84.3%; Score 19.4; DB 29; Length 774;
  Best Local Similarity 95.2%; Pred. No. 4.5e+02;
  Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2  TTAATCATATGCGTTTGGTT 22
      |||||
Db  658 TTAATCATATGCGTTTGGTT 678

RESULT 2
EX134638
LOCUS
DEFINITION
  Danio rerio genomic clone DKEY-91C22, genomic survey sequence.
ACCESSION
  EX134638.1 GI:27965929
KEYWORDS
  GSS.
SOURCE
  Danio rerio (zebrafish)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
  Cypriniformes; Cyprinidae; Danio.
REFERENCE
  1 (bases 1 to 775)
  Humphray S.J., Huckle, E. and Durham, J.L.
  Direct Submission
  Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
  Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
  humquerry@sanger.ac.uk Unpublished
  This sequence was generated from the SP6 end of BAC 91C22. 91C22 is
  part of the Dantoko BAC library created by R. Piasterk and N.V.
  Keygene. Further details:
  http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES
  source
    Location/Qualifiers
      1..775
        /organism="Danio rerio"
        /mol_type="genomic DNA"
        /db_xref="taxon:7955"
        /clone="DKEY-91C22"
        /tissue_type="Testis"
        /note="Vector pIndigoBAC-536"
ORIGIN
  Query Match      82.6%; Score 19; DB 29; Length 775;
  Best Local Similarity 100.0%; Pred. No. 6.6e+02;
  Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1  TTAATCATATGCGTTTGGTT 19
      |||||
Db  267 TTAATCATATGCGTTTGGTT 285

RESULT 3
AZ753174/c
LOCUS
DEFINITION
  RPCI-24-82118.TV RPCI-24 Mus musculus genomic clone RPCI-24-82118,
  genomic survey sequence.
ACCESSION
  AZ753174
VERSION
  AZ753174.1 GI:12538333
KEYWORDS
  GSS.
  Mus musculus (house mouse)
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 616)
  Zhao S., Nierman, M., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
  Teegaye, G., Geer, K., Kroi, M., Shvartsbeyn, A., Gebregorgis, E.,
  Russell, D., de Jong, P. and Fraser, C.M.
  Mouse BAC End Sequences from Library RPCI-24
  Unpublished (1999)
  Other GSSs: RPCI-24-82118.TJ
  Contact: Shaying Zhao
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: szhao@tigr.org
  Clones are derived from the mouse BAC library RPCI-24. For BAC
  library availability, please contact Pieter de Jong
  (pdejong@mail.cho.org). Clones may be purchased from BACPAC
  Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
  page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
  Plate: 82 row: I column: 18
  Seq primer: T7
  Class: BAC ends.
  Location/Qualifiers
    1..616
      /organism="Mus musculus"
      /mol_type="genomic DNA"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="RPCI-24-82118"
      /sex="Male"
      /cell_type="Spleen/Brain"
      /clone_lib="RPCI-24"
      /note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI; The
      RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
      library was cloned in the pTARBAC1 cloning vector at the
      BamHI sites using MboI partially digested male C57BL/6J
      DNA."
ORIGIN
  Query Match      81.7%; Score 18.8; DB 28; Length 616;
  Best Local Similarity 90.9%; Pred. No. 8.1e+02;
  Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  1  TTAATCATATGCGTTTGGTT 22
      |||||
Db  399 TTAATCATATGCGTTTGGTT 378

RESULT 4
CG009365/c
LOCUS
DEFINITION
  ZUAET89TV ZM_3.0.4.0_KB Zea mays genomic clone ZM89Pa0045010,
  genomic survey sequence.
ACCESSION
  CG009365
VERSION
  CG009365.1 GI:33881531
KEYWORDS
  GSS.
SOURCE
  Zea mays
ORGANISM
  Zea mays
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
  clade; Panicoideae; Andropogoneae; Zea.
  1 (bases 1 to 694)
  Whitehead, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
  Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
  Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
  Consortium for Maize Genomics
  Unpublished (2002)
  Other_GSSs: ZUAET89TH
  Contact: Cathy Whitelaw
  Mus musculus (house mouse)
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 616)
  Zhao S., Nierman, M., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
  Teegaye, G., Geer, K., Kroi, M., Shvartsbeyn, A., Gebregorgis, E.,
  Russell, D., de Jong, P. and Fraser, C.M.
  Mouse BAC End Sequences from Library RPCI-24
  Unpublished (1999)
  Other GSSs: RPCI-24-82118.TJ
  Contact: Shaying Zhao
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: szhao@tigr.org
  Clones are derived from the mouse BAC library RPCI-24. For BAC
  library availability, please contact Pieter de Jong
  (pdejong@mail.cho.org). Clones may be purchased from BACPAC
  Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
  page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
  Plate: 82 row: I column: 18
  Seq primer: T7
  Class: BAC ends.
  Location/Qualifiers
    1..616
      /organism="Mus musculus"
      /mol_type="genomic DNA"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="RPCI-24-82118"
      /sex="Male"
      /cell_type="Spleen/Brain"
      /clone_lib="RPCI-24"
      /note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI; The
      RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
      library was cloned in the pTARBAC1 cloning vector at the
      BamHI sites using MboI partially digested male C57BL/6J
      DNA."
ORIGIN
  Query Match      81.7%; Score 18.8; DB 28; Length 616;
  Best Local Similarity 90.9%; Pred. No. 8.1e+02;
  Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  1  TTAATCATATGCGTTTGGTT 22
      |||||
Db  399 TTAATCATATGCGTTTGGTT 378

RESULT 4
CG009365/c
LOCUS
DEFINITION
  ZUAET89TV ZM_3.0.4.0_KB Zea mays genomic clone ZM89Pa0045010,
  genomic survey sequence.
ACCESSION
  CG009365
VERSION
  CG009365.1 GI:33881531
KEYWORDS
  GSS.
SOURCE
  Zea mays
ORGANISM
  Zea mays
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
  clade; Panicoideae; Andropogoneae; Zea.
  1 (bases 1 to 694)
  Whitehead, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
  Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
  Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
  Consortium for Maize Genomics
  Unpublished (2002)
  Other_GSSs: ZUAET89TH
  Contact: Cathy Whitelaw

```

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.

FEATURES
source
Location/Qualifiers
1. .694
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMWBP20045010"
/clone_lib="ZM_3.0_4.0_KB"
/notes="vector: pBCSK-; Site_1: HincII; 3-4 kb 'unfiltered'
genomic DNA library"

ORIGIN
Query Match 81.7%; Score 18.8; DB 29; Length 694;
Best Local Similarity 90.9%; Pred.No.8.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TAATCATGTCGGTTTTGGTTA 23
|||||
Db 275 TAATCTTATGAGTTTTGGTTA 254
|||||

RESULT 5
AQ788301 linear GSS 03-AUG-1999
LOCUS
DEFINITION
HS 3143 A1 G12 MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3143 Col=23 Row=M, genomic survey
sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AQ788301
AQ788301.1 GI:5695925
GSS.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 752)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

REFERENCE
AUTHORS
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

TITLE
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome

JOURNAL
MEDLINE
PUBMED
COMMENT
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
10449764
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: <http://www.htsc.washington.edu>
Plate: 3143 row: M column: 23
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 752.

FEATURES
source
Location/Qualifiers
1. .752
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="plate=3143 Col=23 Row=M"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/notes="Organ: sperm; Vector: pBelcBAC11; BAC Clones in
E-Coli DH10B"

genomic clone 12G15-011, genomic survey sequence.

ACCESSION
AZ045618
VERSION
AZ045618.1 GI:7240056
KEYWORDS
GSS.
SOURCE
Medicago truncatula (barrel medic)

ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE
1 (bases 1 to 475)
AUTHORS
Kim,D., Baek,J., Lim,H., Peng,H., Ellis,L. and Cook,D.R.
TITLE
BAC survey sequencing of Medicago truncatula (2000a)
JOURNAL
Unpublished (2000)
COMMENT
Contact: Cook DR
The Crop Biotechnology Center
Texas A&M University
Department of Plant Pathology and Microbiology, Rm 120 L.F.
Peterson Bldg, College Station, TX 77843-2132, USA
Tel: 409 845 8743
Fax: 409 862 4790
Email: dcook@ppserver.tamu.edu

Other name: BSC-3D-011; date: 3/2/00; Submitted to the Database of
Genome Survey Sequences (GSS) on 03/13/00; More information is
available at <http://chrysis.tamu.edu/medicago>.
Seq primer: pUC-D
Class: BAC subclone.

FEATURES
source

1..475 Location/Qualifiers

/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="12G15-011"
/note="Vector: pUC18; BAC survey sequences were obtained
from sheared BAC DNA subcloned into the SmaI site of
pUC18. The template DNA for sequencing was obtained by PCR
using universal primers. Sequencing reactions were primed
from the pUC-C primer site (CAGGAACAGCTATGACCATGATTACGA)
in the pUC18 polylinker."

ORIGIN

Query Match 80.0%; Score 18.4; DB 28; Length 475;
Best Local Similarity 95.0%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 2 TAATCATATGCGTTTGGT 21
|||||
Db 294 TAATCATATGCGTTTGGT 275

RESULT 11
AZ045411/c
LOCUS
DEFINITION
T234004b shotgun sub-library of BAC clone 10M15 Medicago truncatula
Genomic clone 10M16-004, genomic survey sequence.
ACCESSION
AZ045411
VERSION
AZ045411.1 GI:7239849
KEYWORDS
GSS.
SOURCE
Medicago truncatula (barrel medic)

ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

1 (bases 1 to 491)
AUTHORS
Kim,D., Baek,J., Lim,H., Peng,H., Ellis,L. and Cook,D.R.
TITLE
BAC survey sequencing of Medicago truncatula (2000a)
JOURNAL
Unpublished (2000)
COMMENT
Contact: Cook DR
The Crop Biotechnology Center
Texas A&M University

Department of Plant Pathology and Microbiology, Rm 120 L.F.
Peterson Bldg, College Station, TX 77843-2132, USA
Tel: 409 845 8743
Fax: 409 862 4790
Email: dcook@ppserver.tamu.edu

Other name: BSC-2C-004; date: 3/1/00; Submitted to the Database of
Genome Survey Sequences (GSS) on 03/13/00; More information is
available at <http://chrysis.tamu.edu/medicago>.
Seq primer: pUC-C
Class: BAC subclone.

FEATURES
source

1..491 Location/Qualifiers

/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="10M16-004"
/clone_lib="shotgun sub-library of BAC clone 10M16"
/note="Vector: pUC18; BAC survey sequences were obtained
from sheared BAC DNA subcloned into the SmaI site of
pUC18. The template DNA for sequencing was obtained by PCR
using universal primers. Sequencing reactions were primed
from the pUC-C primer site (CAGGAACAGCTATGACCATGATTACGA)
in the pUC18 polylinker."

ORIGIN

Query Match 80.0%; Score 18.4; DB 28; Length 491;
Best Local Similarity 95.0%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 2 TAATCATATGCGTTTGGT 21
|||||
Db 419 TAATCATATGCGTTTGGT 400

RESULT 12
AQ579547

LOCUS
DEFINITION
T135008b shotgun sub-library of BAC clone 10M16 Medicago truncatula
genomic clone 10-M-16-C-008, genomic survey sequence.
ACCESSION
AQ579547
VERSION
AQ579547.1 GI:4979622
KEYWORDS
GSS.
SOURCE
Medicago truncatula (barrel medic)

ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

1 (bases 1 to 574)
AUTHORS
Kim,D., Peng,H., Ellis,L. and Cook,D.R.
TITLE
BAC survey sequencing of Medicago truncatula
JOURNAL
Unpublished (1999)
COMMENT
Contact: Cook DR
The Crop Biotechnology Center
Texas A&M University
Department of Plant Pathology and Microbiology, Rm 120 L.F.
Peterson Bldg, College Station, TX 77843-2132, USA
Tel: 409 845 8743
Fax: 409 862 4790
Email: dcook@ppserver.tamu.edu

Other name: BSC-2-08; date: 3/3/99; Submitted to the Database of
Genome Survey Sequences (GSS) on 06/01/99; More information is
available at <http://chrysis.tamu.edu/medicago>.
Seq primer: pUC-C
Class: BAC subclone.

FEATURES
source

1..574 Location/Qualifiers

/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="10-M-16-C-008"

/clone lib="shotgun sub-library of BAC clone 10M16"
 /note="Vector: pUC18; BAC survey sequences were obtained
 from sheared BAC DNA subcloned into the SmaI site of
 pUC18. The template DNA for sequencing was obtained by PCR
 using universal primers. Sequencing reactions were primed
 from the pUC-C primer site (CAGGAAACGCTATGACCATGATTACGA)
 in the pUC18 polylinker."

ORIGIN

Query Match 80.0%; Score 18.4; DB 28; Length 574;
 Best Local Similarity 95.0%; Pred. No. 1.2e+03;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TAATCATATGCGTTTGGT 21
 |||||
 Db 268 TAATCATATGCGTTTGGT 287

RESULT 13

AQ579574/c
 LOCUS T135037b shotgun sub-library of BAC clone 10M16 Medicago truncatula
 DEFINITION genomic clone 10-M-16-C-037, genomic survey sequence.
 ACCESSION AQ579574
 VERSION AQ579574.1 GI:4979649
 KEYWORDS GSS.
 SOURCE Medicago truncatula (barrel medic)

ORGANISM

Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliopsida; eudicotyledons; Core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.

REFERENCE

1 (bases 1 to 655)
 Kim, D., Peng, H., Ellis, L., and Cook, D.R.
 BAC survey sequencing of Medicago truncatula
 Unpublished (1999)

JOURNAL

COMMENT

Contact: Cook DR
 The Crop Biotechnology Center
 Texas A&M University
 Department of Plant Pathology and Microbiology, Rm 120 L.F.
 Peterson Bldg, College Station, TX 77843-2132, USA
 Tel: 409 845 8743
 Fax: 409 862 4790

Email: dcooke@ppserver.tamu.edu
 Other name: BSC-2-37; date: 3/3/99; Submitted to the Database of
 Genome Survey Sequences (GSS) on 06/01/99; More information is
 available at 'http://chrysis.tamu.edu/medicago'.

Seq primer: pUC-C

Class: BAC subclone.

FEATURES

source

1..655
 Location/Qualifiers
 /organism="Medicago truncatula"
 /mol_type="genomic DNA"
 /culti_var="genotype AL7"
 /db_xref="taxon:3880"
 /clone="10-M-16-C-037"
 /clone_lib="shotgun sub-library of BAC clone 10M16"
 /note="Vector: pUC18; BAC survey sequences were obtained
 from sheared BAC DNA subcloned into the SmaI site of
 pUC18. The template DNA for sequencing was obtained by PCR
 using universal primers. Sequencing reactions were primed
 from the pUC-C primer site (CAGGAAACGCTATGACCATGATTACGA)
 in the pUC18 polylinker."

ORIGIN

Query Match 80.0%; Score 18.4; DB 28; Length 655;
 Best Local Similarity 95.0%; Pred. No. 1.2e+03;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TAATCATATGCGTTTGGT 21
 |||||
 Db 386 TAATCATATGCGTTTGGT 367

RESULT 14

BM164586

LOCUS

DEFINITION

BM164586

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BM164586 309 bp mRNA linear EST 04-DEC-2001
 EST567109 PyBS Plasmodium yoelii yoelii cDNA clone PYCLW28 5' end,
 mRNA sequence.
 BM164586
 BM164586.1 GI:17310267
 EST.
 Plasmodium yoelii yoelii
 Plasmodium yoelii yoelii
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 309)
 Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B.,
 Fraser, C.M., and Carucci, D.J.
 Plasmodium yoelii EST project at TIGR
 Unpublished (2001)
 Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADP.

FEATURES

source

1..309
 Location/Qualifiers
 /organism="Plasmodium yoelii yoelii"
 /mol_type="mRNA"
 /strain="17XL"
 /sub_species="yoelii"
 /db_xref="taxon:73239"
 /clone="PYCLW28"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /clone_lib="PyBS"
 /note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
 collected from BALB/cBYJ mice infected with Py17XL
 parasites, and leukocytes removed by passage over
 microcrystalline cellulose columns. Total RNA was
 isolated using the guanidinium isothiocyanate method, and
 mRNA isolated using oligo(dT)-cellulose chromatography.
 First strand cDNA synthesis was completed using a 50-base
 primer and reverse transcriptase in the presence of
 5-methyl dCTP. After second strand synthesis, uneven
 termini were treated with Pfu DNA polymerase and EcoRI
 adaptors ligated to the blunt ends. The sample was cleaved
 with XhoI and separated on a Sephacryl S-500 column.
 Size-fractionated cDNA was precipitated and ligated to
 HybrizAP arms directionally using EcoRI-XhoI cleaved arms.
 After packaging, the phagemid vector (PAD-GAL4) was
 excised from the HybrizAP vector and plasmid DNA
 isolated."

ORIGIN

Query Match 79.1%; Score 18.2; DB 12; Length 309;
 Best Local Similarity 87.0%; Pred. No. 1.4e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TTAATCATATGCGTTTGGTTA 23
 |||||
 Db 90 TTAATCATATGCGTTTGGTTA 112

RESULT 15

AL1738836

LOCUS

DEFINITION

AL1738836

ACCESSION

AL1738836 343 bp mRNA linear EST 18-JUN-1999
 tr28911.x1 NCI_CGAP_Ov23 Homo sapiens cDNA clone IMAGE:2219684 3',
 mRNA sequence.
 AL1738836

```

VERSION      AI738836.1  GI:5100817
KEYWORDS     EST
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 343)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
              Emmert-Buck, M.D., Ph.D.
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              www-bio.llnl.gov/bbrp/image/image.html
              Seq primer: -400P from Gibco
              High quality sequence stop: 337.

FEATURES
  source
    1..343
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:2219684"
      /tissue_type="tumor, 5 pooled (see description)"
      /lab_host="DH10B"
      /clone_lib="NCI_CGAP_Ov23"
      /notes="Organ: ovary; Vector: pCMV-SPORT6; Site: 1: SalI;
      Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
      Average insert size 1.35 Kb. Tumor types include: mixed
      Mullerian tumor, papillary serous, clear cell, spindle
      cell. All are primary tumors, metastasis positive. Life
      Technologies catalog #: 11534-013"

ORIGIN
Query Match      79.1%; Score 18.2; DB 9; Length 343;
Best Local Similarity 87.0%; Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1  TTATCATATGCGCTTTTGCTTA 23
Db      78  TTATCAGATGCGCTTTTAGTTA 100

Search completed: May 26, 2004, 22:27:08
Job time : 1597.59 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 15:22:28 ; Search time 178.526 Seconds
(without alignments)
594.900 Million cell updates/sec

Title: US-10-676-299-9

Perfect score: 25
Sequence: i ttaagtcataatgtttttgactta 25

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04:*

- 1: geneseqn1980s:*
- 2: geneseqn1990s:*
- 3: geneseqn2000s:*
- 4: geneseqn2001as:*
- 5: geneseqn2001bs:*
- 6: geneseqn2002s:*
- 7: geneseqn2003as:*
- 8: geneseqn2003bs:*
- 9: geneseqn2003cs:*
- 10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	25	100.0	25	8	ACD28591 E. coli A
C 2	25	100.0	25	8	ACD28590 B. coli A
C 3	25	100.0	40	8	ACD28587 E. coli A
C 4	25	100.0	42	8	ACD28586 E. coli A
C 5	25	100.0	401	7	ACA15469
C 6	25	100.0	2182	5	AAS82555 DNA encod
C 7	20.2	80.8	2000	6	AB215842 Arabidops
C 8	20.2	80.8	5641	6	ABL33397 Human inn
C 9	19.8	79.2	62782	8	AAD58282 Human tum
C 10	19.8	79.2	62782	8	AAD58281
C 11	19.8	79.2	94191	9	ADE11169_3
C 12	19.8	79.2	226475	8	AAD58279
C 13	19.2	76.8	1293	3	AAC36413
C 14	18.8	75.2	110000	6	ABA92787_2
C 15	18.6	74.4	561	5	ABA13700
C 16	18.6	74.4	668	5	ABA19908
C 17	18.6	74.4	668	5	ABA20194
C 18	18.6	74.4	879	5	ABA20198
C 19	18.6	74.4	1772	5	ABA19910
C 20	18.6	74.4	1772	5	ABA20197
C 21	18.6	74.4	1772	5	ABA18947
C 22	18.6	74.4	1772	5	ABA20195
C 23	18.6	74.4	1772	5	ABA19911

C 24	18.6	74.4	1772	5	ABA18949	Human ner
C 25	18.6	74.4	2585	5	ABA19906	Human ner
C 26	18.6	74.4	2585	5	ABA18948	Human ner
C 27	18.6	74.4	2585	5	ABA20196	Human ner
C 28	18.6	74.4	10279	6	ABL33591	Human inn
C 29	18.6	74.4	10279	6	ABL92277	Chemical
C 30	18.6	74.4	10279	6	AAD22328	Chemical
C 31	18.6	74.4	10311	4	AAK84424	Human inn
C 32	18.6	74.4	10311	4	AAK84423	Human inn
C 33	18.6	74.4	73334	6	ABL34124	Human inn
C 34	18.6	74.4	73334	6	ABL92318	Chemical
C 35	18.4	73.6	10891	6	ABL32465	Human inn
C 36	18.2	72.8	583	5	AAS33559	Human DNA
C 37	18.2	72.8	891	5	AAS34560	Human DNA
C 38	18.2	72.8	1486	3	AAZ97081	Human sec
C 39	18.2	72.8	1486	8	ACH66710	Novel hum
C 40	18.2	72.8	2121	6	AAI43413	A thalian
C 41	18.2	72.8	10467	6	ABL49301	Human pol
C 42	18.2	72.8	10872	4	AAI03182	Human rep
C 43	18.2	72.8	53905	7	ACF30939	Rice cult
C 44	18.2	72.8	76363	7	ACF30938	Rice cult
C 45	18.2	72.8	349881	9	ADC86642	Human GPC

ALIGNMENTS

RESULT 1
ACD28591/c
ID ACD28591 standard; DNA; 25 BP.
XX
AC ACD28591;
XX
DT 10-OCT-2003 (first entry)
XX
DE E. coli ArsrR binding oligonucleotide CHROMSIB.
XX
KW Arsr; arsenic resistance operon; biosensor; ss; arsenic.
OS Escherichia coli.
XX
PN US2003096275-A1.
XX
PD 22-MAY-2003.
XX
PF 15-AUG-2002; 2002US-00222952.
XX
PR 20-AUG-2001; 2001US-0313714P.
XX
(LAIN/) LAING L G.
XX
Laing LG;
XX
WPI: 2003-576876/54.
XX
New system comprising isolated protein and nucleic acid, and a detection system that indicates a change in binding of the protein to the nucleic acid in the presence of the analyte, useful for detecting the presence of analyte in a sample.
XX
Claim 35; Page 15; 36pp; English.
XX
The invention relates to a new system (biosensor) for detecting the presence of analyte in a sample comprising: (1) an isolated protein that specifically binds the analyte; (2) an isolated nucleic acid containing a specific binding sequence that is bound specifically by the protein; and (3) a detection system that indicates a change in binding of the protein to the nucleic acid in the presence of the analyte. Also included are a biosensor device for detecting the presence of an analyte in a sample, detecting the presence of an analyte in a sample, and an ArsrR (encoded by part of the arsenic resistance operon of E. coli) protein comprising an amino acid sequence that is at least 90% identical to amino acids 1-97 of the ArsrR sequence appearing as ABU63440 binding to a nucleic acid

CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the bottom strand of an oligonucleotide which binds
 CC to a chromosomally expressed ArsrR protein and is used in the biosensor of
 CC the invention

XX SQ Sequence 25 BP; 13 A; 3 C; 2 G; 7 T; 0 U; 0 Other;

Query Match 100.0%; Score 25; DB 8; Length 25;

Best Local Similarity 100.0%; Pred. No. 1.6; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTAAGTCATATATGTTTTGACTTA 25
 |||||
 Db 25 TTAAGTCATATATGTTTTGACTTA 1

RESULT 2

ACD28590

ID ACD28590 standard; DNA; 25 BP.

XX AC ACD28590;

XX DT 10-OCT-2003 (first entry)

XX DE E. coli ArsrR binding oligonucleotide CHROMS1T.

XX KW ArsrR; arsenic resistance operon; biosensor; ss; arsenic.

XX OS Escherichia coli.

XX PN US2003096275-A1.

XX PD 22-MAY-2003.

XX PF 15-AUG-2002; 2002US-00222952.

XX PR 20-AUG-2001; 2001US-0313714P.

XX PA (LAIN/) LAING L G.

XX PI Laing LG;

XX WPI; 2003-576876/54.

XX FT New system comprising isolated protein and nucleic acid, and a detection
 PT system that indicates a change in binding of the protein to the nucleic
 PT acid in the presence of the analyte, useful for detecting the presence of
 PT analyte in a sample.

XX PS Claim 35; Page 15; 36pp; English.

XX CC The invention relates to a new system (biosensor) for detecting the
 CC presence of analyte in a sample comprising: (1) an isolated protein that
 CC specifically binds the analyte; (2) an isolated nucleic acid containing a
 CC specific binding sequence that is bound specifically by the protein; and
 CC (3) a detection system that indicates a change in binding of the protein
 CC to the nucleic acid in the presence of the analyte. Also included are a
 CC biosensor device for detecting the presence of an analyte in a sample,
 CC detecting the presence of an analyte in a sample, and an ArsrR (encoded by
 CC part of the arsenic resistance operon of E. coli) protein comprising an
 CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
 CC the ArsrR sequence appearing as ABU63440 binding to a nucleic acid
 CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the top strand of an oligonucleotide which binds to a
 CC chromosomally expressed ArsrR protein and is used in the biosensor of the
 CC invention

XX SQ Sequence 25 BP; 7 A; 2 C; 3 G; 13 T; 0 U; 0 Other;

Query Match

Best Local Similarity 100.0%; Score 25; DB 8; Length 25;

100.0%; Pred. No. 1.6;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTAAGTCATATATGTTTTGACTTA 25

|||||
 Db 1 TTAAGTCATATATGTTTTGACTTA 25

RESULT 3

ACD28587/c

ID ACD28587 standard; DNA; 40 BP.

XX AC ACD28587;

XX DT 10-OCT-2003 (first entry)

XX DE E. coli ArsrR binding oligonucleotide CHROML1B.

XX KW ArsrR; arsenic resistance operon; biosensor; ss; arsenic.

XX OS Escherichia coli.

XX PN US2003096275-A1.

XX PD 22-MAY-2003.

XX PF 15-AUG-2002; 2002US-00222952.

XX PR 20-AUG-2001; 2001US-0313714P.

XX PA (LAIN/) LAING L G.

XX PI Laing LG;

XX WPI; 2003-576876/54.

XX FT New system comprising isolated protein and nucleic acid, and a detection
 PT system that indicates a change in binding of the protein to the nucleic
 PT acid in the presence of the analyte, useful for detecting the presence of
 PT analyte in a sample.

XX PS Claim 35; Page 15; 36pp; English.

XX CC The invention relates to a new system (biosensor) for detecting the
 CC presence of analyte in a sample comprising: (1) an isolated protein that
 CC specifically binds the analyte; (2) an isolated nucleic acid containing a
 CC specific binding sequence that is bound specifically by the protein; and
 CC (3) a detection system that indicates a change in binding of the protein
 CC to the nucleic acid in the presence of the analyte. Also included are a
 CC biosensor device for detecting the presence of an analyte in a sample,
 CC detecting the presence of an analyte in a sample, and an ArsrR (encoded by
 CC part of the arsenic resistance operon of E. coli) protein comprising an
 CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
 CC the ArsrR sequence appearing as ABU63440 binding to a nucleic acid
 CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the bottom strand of an oligonucleotide which binds
 CC to a chromosomally expressed ArsrR protein and is used in the biosensor of
 CC the invention

XX SQ Sequence 40 BP; 17 A; 5 C; 7 G; 11 T; 0 U; 0 Other;

Query Match

Best Local Similarity 100.0%; Score 25; DB 8; Length 40;

100.0%; Pred. No. 1.6;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTAAGTCATATATGTTTTGACTTA 25

|||||
 Db 25 TTAAGTCATATATGTTTTGACTTA 1

RESULT 4

ACD28586

ID ACD28586 standard; DNA; 42 BP.

XX AC ACD28586;
 XX DT 10-OCT-2003 (first entry)
 XX DE E. coli ArsR binding oligonucleotide CHROMELIT.
 XX KW ArsR; arsenic resistance operon; biosensor; ss; arsenic.
 XX OS Escherichia coli.
 XX PN US2003096275-A1.
 XX PD 22-MAY-2003.
 XX PF 15-AUG-2002; 2002US-00222952.
 XX PR 20-AUG-2001; 2001US-0313714P.
 XX PA (LAIN/) LAING L G.
 XX PI Laing LG;
 XX DR WPI; 2003-576876/54.
 XX PT New system comprising isolated protein and nucleic acid, and a detection
 PT system that indicates a change in binding of the protein to the nucleic
 PT acid in the presence of the analyte, useful for detecting the presence of
 XX analyte in a sample.
 XX FS Claim 35; Page 15; 36pp; English.
 XX CC The invention relates to a new system (biosensor) for detecting the
 CC presence of analyte in a sample comprising: (1) an isolated protein that
 CC specifically binds the analyte; (2) an isolated nucleic acid containing a
 CC specific binding sequence that is bound specifically by the protein; and
 CC (3) a detection system that indicates a change in binding of the protein
 CC to the nucleic acid in the presence of the analyte. Also included are a
 CC biosensor device for detecting the presence of an analyte in a sample,
 CC detecting the presence of an analyte in a sample, and an ArsR (encoded by
 CC part of the arsenic resistance operon of E. coli) protein comprising an
 CC amino acid sequence that is at least 90% identical to amino acids 1-97 of
 CC the ArsR sequence appearing as A063440 binding to a nucleic acid
 CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the top strand of an oligonucleotide which binds to a
 CC chromosomally expressed ArsR protein and is used in the biosensor of the
 XX invention
 XX SQ Sequence 42 BP; 11 A; 8 C; 5 G; 18 T; 0 U; 0 Other;
 Query Match 100.0%; Score 25; DB 8; Length 42;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTAAGTCATATATGTTTGGACTTA 25
 Db 18 TTAAGTCATATATGTTTGGACTTA 42
 RESULT 5
 ACA15469/c
 ID ACA15469 standard; DNA; 401 BP.
 XX AC ACA15469;
 XX DT 27-OCT-2003 (revised)
 XX DT 19-JUN-2003 (first entry)
 XX DE Prokaryotic essential gene antisense oligonucleotide #3339.
 XX KW Antisense; ss; prokaryotic essential gene; cell proliferation;
 KW drug design.

XX Archaea.
 XX OS WC200277183-A2.
 XX PN 03-OCT-2002.
 XX PD 21-MAR-2002; 2002WO-US009107.
 XX PF 21-MAR-2001; 2001US-00815242.
 XX PR 06-SEP-2001; 2001US-00948993.
 XX PR 25-OCT-2001; 2001US-0342923P.
 XX PR 08-FEB-2002; 2002US-00072851.
 XX PR 06-MAR-2002; 2002US-0362699P.
 XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Mall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX DR WPI; 2003-029926/02.
 XX PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX FS Claim 1; SEQ ID NO 339; 1766pp; English.
 XX CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC the gene product or that has an activity against a biological pathway;
 CC the gene product or that has an activity against a biological pathway;
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than S. aureus, S. typhimurium,
 CC K. pneumoniae or P. aeruginosa. The present sequence is one of the 6213
 CC antisense sequences of the invention. Note: the sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to
 CC standardise OS field)
 XX SQ Sequence 401 BP; 113 A; 68 C; 101 G; 119 T; 0 U; 0 Other;
 Query Match 100.0%; Score 25; DB 7; Length 401;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTAAGTCATATATGTTTGGACTTA 25
 Db 128 TTAAGTCATATATGTTTGGACTTA 104
 RESULT 6

```

AAS82555
ID AAS82555 standard; cDNA; 2182 BP.
XX AC
XX AC
XX AAS82555;
XX DT
XX DT
XX 13-FEB-2002 (first entry)
XX DE
XX DNA encoding novel human diagnostic protein #18359.
XX KW
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS
XX Homo sapiens.
XX PW
XX WO200175067-A2.
XX PD
XX 11-OCT-2001.
XX PF
XX 30-MAR-2001; 2001WO-US008631.
XX PR
XX 31-MAR-2000; 2000US-00540217.
XX PR
XX 23-AUG-2000; 2000US-00549167.
XX PA
XX (HYSE-) HYSEQ INC.
XX PI
XX Dmanac RT, Liu C, Tang YT;
XX DR
XX WPI; 2001-639362/73.
XX DR
XX P-PSDB; ABG18368.
XX PT
XX New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensic, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX XX
XX Claim 1; SEQ ID NO 18359; 103pp; English.
XX CC
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
XX CC coding sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ
XX Sequence 2182 BP; 516 A; 569 C; 582 G; 515 T; 0 U; 0 Other;
XX ID ABL33397 standard; DNA; 5641 BP.
XX AC
XX ABL33397;
XX AC
XX ABL33397;
XX DT
XX 26-MAR-2002 (first entry)
XX XX
XX Human immune system associated gene SEQ ID NO: 1370.
XX DE
XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX KW antiarteriosclerotic; antianaemic; cytosolic; neutropenic;
XX KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX KW antineuritic; antiarthritic; antidiabetic; antipsoriatic;
XX KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX ID ABL33397 standard; cDNA; 2182 BP.
XX AC
XX AAS82555;
XX DT
XX 13-FEB-2002 (first entry)
XX DE
XX DNA encoding novel human diagnostic protein #18359.
XX KW
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS
XX Homo sapiens.
XX PW
XX WO200175067-A2.
XX PD
XX 11-OCT-2001.
XX PF
XX 30-MAR-2001; 2001WO-US008631.
XX PR
XX 31-MAR-2000; 2000US-00540217.
XX PR
XX 23-AUG-2000; 2000US-00549167.
XX PA
XX (HYSE-) HYSEQ INC.
XX PI
XX Dmanac RT, Liu C, Tang YT;
XX DR
XX WPI; 2001-639362/73.
XX DR
XX P-PSDB; ABG18368.
XX PT
XX New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensic, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX XX
XX Claim 1; SEQ ID NO 18359; 103pp; English.
XX CC
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
XX CC coding sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ
XX Sequence 2182 BP; 516 A; 569 C; 582 G; 515 T; 0 U; 0 Other;
XX ID ABL33397 standard; DNA; 5641 BP.
XX AC
XX ABL33397;
XX AC
XX ABL33397;
XX DT
XX 26-MAR-2002 (first entry)
XX XX
XX Human immune system associated gene SEQ ID NO: 1370.
XX DE
XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX KW antiarteriosclerotic; antianaemic; cytosolic; neutropenic;
XX KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX KW antineuritic; antiarthritic; antidiabetic; antipsoriatic;
XX KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

```

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.
 XX Homo sapiens.
 OS WO200200928-A2.
 PN
 XX 03-JAN-2002.
 PD
 XX 02-JUL-2001; 2001WO-BP007537.
 FF
 XX 30-JUN-2000; 2000DE-01032529.
 PR
 XX 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPIC-) EPICENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-130909/17.
 XX
 XX Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 FT
 XX
 XX Claim 1; SEQ ID NO 1370; 32pp + Sequence Listing; German.
 PS
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC which are used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX
 XX Sequence 5641 BP; 2137 A; 34 C; 1132 G; 2338 T; 0 U; 0 Other;
 SQ
 Query Match 80.8%; Score 20.2; DB 6; Length 5641;
 Best Local Similarity 88.0%; Pred. No. 1.3e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 TTAAGTCATATATGTTTGGACTTA 25
 DB 1512 TTAAGTCATATATATTTTGGATTA 1536
 RESULT 9
 AAD58282/c
 ID AAD58282 standard; DNA; 62782 BP.
 XX
 AC AAD58282;
 XX
 XX 20-NOV-2003 (first entry)
 DT
 XX Human tumour suppressor gene, Lmt intron 3 DNA.
 DE
 XX Tumour suppressor gene; Lmt; cancer; therapy; cytostatic; human; ds.
 KW
 XX Homo sapiens.
 OS
 XX WO2003066869-A1.
 PN
 XX 14-AUG-2003.
 PD
 XX 07-FEB-2003; 2003WO-AU000126.
 PF
 XX 07-FEB-2002; 2002AU-00000371.
 PR
 XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 PA
 XX Cook WD, Mccaw BJ;
 PI WPI; 2003-646311/61.
 XX
 XX New nucleic acid molecule, useful for screening a subject for the
 PT presence of an aberration in a gene encoding an LMT.
 FT
 XX Claim 10; Page 314-333; 373pp; English.
 PS
 XX The invention relates to novel tumour suppressor gene, referred to as
 CC Lmt. The invention also relates to the field of cancer therapy and cancer
 CC diagnostics. The nucleic acid molecule is useful for screening a subject
 CC for the presence of an aberration in a gene encoding an LMT. The present
 CC sequence is human Lmt intron 2 DNA
 XX
 XX Sequence 62782 BP; 15441 A; 9614 C; 9374 G; 14439 T; 0 U; 13914 Other;
 SQ
 Query Match 79.2%; Score 19.8; DB 8; Length 62782;
 Best Local Similarity 91.3%; Pred. No. 1.9e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 TAAAGTCATATATGTTTGGACTT 24
 DB 28459 TAAAGTCATATATGTTTGGACAT 28437
 RESULT 10
 AAD58281/c
 ID AAD58281 standard; DNA; 62782 BP.
 XX
 AC AAD58281;
 XX
 XX 20-NOV-2003 (first entry)
 DT
 XX Human tumour suppressor gene, Lmt intron 2 DNA.
 DE
 XX Tumour suppressor gene; Lmt; cancer; therapy; cytostatic; human; ds.
 KW
 XX Homo sapiens.
 OS
 XX WO2003066869-A1.
 PN
 XX 14-AUG-2003.
 PD
 XX 07-FEB-2003; 2003WO-AU000126.
 PF
 XX 07-FEB-2002; 2002AU-00000371.
 PR
 XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 PA
 XX Cook WD, Mccaw BJ;
 PI WPI; 2003-646311/61.
 XX
 XX New nucleic acid molecule, useful for screening a subject for the
 PT presence of an aberration in a gene encoding an LMT.
 FT
 XX Claim 10; Page 314-333; 373pp; English.
 PS
 XX The invention relates to novel tumour suppressor gene, referred to as
 CC Lmt. The invention also relates to the field of cancer therapy and cancer
 CC diagnostics. The nucleic acid molecule is useful for screening a subject
 CC for the presence of an aberration in a gene encoding an LMT. The present
 CC sequence is human Lmt intron 2 DNA
 XX
 XX Sequence 62782 BP; 15441 A; 9614 C; 9374 G; 14439 T; 0 U; 13914 Other;
 SQ
 Query Match 79.2%; Score 19.8; DB 8; Length 62782;
 Best Local Similarity 91.3%; Pred. No. 1.9e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 TAAAGTCATATATGTTTGGACTT 24
 DB 28459 TAAAGTCATATATGTTTGGACAT 28437

XX New nucleic acid molecule, useful for screening a subject for the
 PT presence of an aberration in a gene encoding an LMT.
 FT
 XX Claim 10; Page 333-358; 373pp; English.
 PS
 XX The invention relates to novel tumour suppressor gene, referred to as
 CC Lmt. The invention also relates to the field of cancer therapy and cancer
 CC diagnostics. The nucleic acid molecule is useful for screening a subject
 CC for the presence of an aberration in a gene encoding an LMT. The present
 CC sequence is human Lmt intron 3 DNA
 XX
 XX Sequence 62782 BP; 15441 A; 9614 C; 9374 G; 14439 T; 0 U; 13914 Other;
 SQ

Query Match 79.2%; Score 19.8; DB 8; Length 62782;
 Best Local Similarity 91.3%; Pred. No. 1.9e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TAAAGTCATATATGTTTGGACTT 24
 DB 28459 TAAAGTCATATATGTTTGGACAT 28437

RESULT 10

AAD58281/c
 ID AAD58281 standard; DNA; 62782 BP.

XX
 AC AAD58281;

XX 20-NOV-2003 (first entry)

XX Human tumour suppressor gene, Lmt intron 2 DNA.

XX Tumour suppressor gene; Lmt; cancer; therapy; cytostatic; human; ds.

XX Homo sapiens.

XX WO2003066869-A1.

XX 14-AUG-2003.

XX 07-FEB-2003; 2003WO-AU000126.

XX 07-FEB-2002; 2002AU-00000371.

XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.

XX Cook WD, Mccaw BJ;

XX WPI; 2003-646311/61.

XX New nucleic acid molecule, useful for screening a subject for the
 PT presence of an aberration in a gene encoding an LMT.
 FT

XX Claim 10; Page 314-333; 373pp; English.

XX The invention relates to novel tumour suppressor gene, referred to as
 CC Lmt. The invention also relates to the field of cancer therapy and cancer
 CC diagnostics. The nucleic acid molecule is useful for screening a subject
 CC for the presence of an aberration in a gene encoding an LMT. The present
 CC sequence is human Lmt intron 2 DNA
 XX

XX Sequence 62782 BP; 15441 A; 9614 C; 9374 G; 14439 T; 0 U; 13914 Other;

Query Match 79.2%; Score 19.8; DB 8; Length 62782;

Best Local Similarity 91.3%; Pred. No. 1.9e+02;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TAAAGTCATATATGTTTGGACTT 24

DB 28459 TAAAGTCATATATGTTTGGACAT 28437

```

RESULT 11
AD11169 3/c
Continuation (4 of 4) of AD11169 from base 300001 (Human transporter protein encoding s
WP Sequence split into 4 fragments LOCUS AD11169 Accession AD11169
WP Fragment Name Begin End
WP AD11169_0 1 110000
WP AD11169_1 100001 210000
WP AD11169_2 200001 310000
WP AD11169_3 300001 394191

Query Match 79.2%; Score 19.8; DB 9; Length 94191;
Best Local Similarity 91.3%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 3 AAGTCATATATGTTTGGACTTA 25
Db 35207 AATGCATATATGTTTGGACTTA 35185

RESULT 12
AAD58279/c
ID AAD58279 standard; DNA; 226475 BP.
XX
AC AAD58279;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human tumour suppressor gene, lmt reverse complement DNA.
XX
KW Tumour suppressor gene; lmt; cancer; therapy; cytostatic; human; ds.
XX
OS Homo sapiens.
XX
PN WO2003066869-A1.
XX
PD 14-AUG-2003.
XX
PF 07-FEB-2003; 2003WO-AU000126.
XX
PR 07-FEB-2002; 2002AU-00090371.
XX
PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
XX
PI Cook WD, Mccaw BJ;
XX
DR WPI; 2003-646311/61.
XX
PT New nucleic acid molecule, useful for screening a subject for the
PT presence of an aberration in a gene encoding an lmt.
XX
PS Claim 10; Page 233-299; 373pp; English.
XX
CC The invention relates to novel tumour suppressor gene, referred to as
CC lmt. The invention also relates to the field of cancer therapy and cancer
CC diagnostics. The nucleic acid molecule is useful for screening a subject
CC for the presence of an aberration in a gene encoding an lmt. The present
CC sequence is human lmt reverse complement DNA
XX
SQ Sequence 226475 BP; 61024 A; 41761 C; 40916 G; 57494 T; 0 U; 25280 Other;

Query Match 79.2%; Score 19.8; DB 8; Length 226475;
Best Local Similarity 91.3%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 2 TAAGTCATATATGTTTGGACTT 24
Db 140950 TAAGTCATATATGTTTGGACTT 140928

RESULT 13
AAC36413
ID AAC36413 standard; DNA; 1293 BP.
XX

```

```

AAC36413;
AC
XX 17-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 13710.
DE
XX Arabidopsis thaliana.
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway; metabolic pathway;
XX promoter; termination sequence; ss.
XX Arabidopsis thaliana.
OS
XX Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PP 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.

```

```
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 24-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140921P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 08-JUL-1999; 99US-0142820P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143342P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 18-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144431P.
PR 19-JUL-1999; 99US-0144433P.
PR 19-JUL-1999; 99US-0144433P.
PR 19-JUL-1999; 99US-0144433P.
PR 19-JUL-1999; 99US-0144433P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145313P.
PR 27-JUL-1999; 99US-0145318P.
PR 27-JUL-1999; 99US-0145319P.
PR 28-JUL-1999; 99US-0145351P.
PR 28-JUL-1999; 99US-0145351P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 15-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149920P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
```

```
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 12-SEP-1999; 99US-0153758P.
PR 13-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156598P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-015753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0158294P.
PR 13-OCT-1999; 99US-0158295P.
PR 14-OCT-1999; 99US-0158329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
```

Query Match 76.8%; Score 19.2; DB 3; Length 1293;
Best Local Similarity 87.5%; Pred. No. 3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 TTAAAGTCATATATGTTTGACTT 24
    ||||| ||||| ||||| |||||
Db 364 TTAAAGTCATATATGTTTGACTT 387
```

RESULT 14

ABA92787_2/c
Continuation {3 of 7} of ABA92787 from base 200001 (Buchnera sp. genomic DNA SEQ ID NO:
WP Sequence split into 7 fragments LOCUS ABA92787 Accession ABA92787

WP	Fragment Name	Begin	End
WP	ABA92787_0	1	110000
WP	ABA92787_1	100001	210000
WP	ABA92787_2	200001	310000
WP	ABA92787_3	300001	410000
WP	ABA92787_4	400001	510000
WP	ABA92787_5	500001	610000

```
WP ABA92787_6 600001 640681
Query Match 75.2%; Score 18.8; DB 6; Length 110000;
Best Local Similarity 90.9%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAGTCATATATGTTTTTGACIT 24
   |||||
Db 23338 AATCATATATTTTTTGACIT 23317

RESULT 15
ID ABA13700/c
XX ABA13700 standard; cDNA; 561 BP.
AC ABA13700;
XX
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 2707.
XX
XX Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antiscikling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
XX WO200159063-A2.
XX
PD 16-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001334.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 18-AUG-2000; 2000US-0225759P.
PR 22-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 23-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 05-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0249287P.
PR 17-NOV-2000; 2000US-0249288P.
PR 17-NOV-2000; 2000US-0249289P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
```

PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249219P.
PR 17-NOV-2000; 2000US-0249220P.
PR 17-NOV-2000; 2000US-0249221P.
PR 17-NOV-2000; 2000US-0249222P.
PR 17-NOV-2000; 2000US-0249223P.
PR 17-NOV-2000; 2000US-0249224P.
PR 17-NOV-2000; 2000US-0249225P.
PR 17-NOV-2000; 2000US-0249226P.
PR 17-NOV-2000; 2000US-0249227P.
PR 17-NOV-2000; 2000US-0249228P.
PR 17-NOV-2000; 2000US-0249229P.
PR 17-NOV-2000; 2000US-0249230P.
PR 01-DEC-2000; 2000US-0250391P.
PR 01-DEC-2000; 2000US-0251160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-02559678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-541565/60.
XX P-PSDE; ABB17374.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system cancers
PT and metastases.
PT
XX
PS Claim 1; SEQ ID NO 2707; 1701pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (AB11004-ABA21534) and proteins
CC (AB114678-AB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 561 BP; 217 A; 96 C; 84 G; 155 T; 0 U; 9 Other;

Query Match 74.4%; Score 18.6; DB 5; Length 561;
Best Local Similarity 84.0%; Pred. No. 5.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 TTAAGTCATATGCTTTTGGACTTA 25
Db 407 TTAAGACATGTTGTTTGGACTCA 383

Search completed: May 26, 2004, 17:50:16
Job time : 182.526 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 16:18:00 ; Search time 39.8504 Seconds
(without alignments)
348.146 Million cell updates/sec

Title: US-10-676-299-9
Perfect score: 25
Sequence: 1 ttaagtcataatgttttgactta 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/prodata/2/ina/5A COMB.seq.*
2: /cgn2_6/prodata/2/ina/5B COMB.seq.*
3: /cgn2_6/prodata/2/ina/6A COMB.seq.*
4: /cgn2_6/prodata/2/ina/6B COMB.seq.*
5: /cgn2_6/prodata/2/ina/PCFUS COMB.seq.*
6: /cgn2_6/prodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18.8	75.2	640681	4	US-09-790-988-1
C 2	18.2	72.8	1486	4	US-09-461-325-73
C 3	18.2	72.8	1486	4	US-10-012-542-73
C 4	18.2	72.8	10457	4	US-10-204-708-1
C 5	17.8	71.2	133303	4	US-09-497-855A-37
C 6	17.8	71.2	133303	4	US-09-497-855A-44
C 7	17.6	70.4	711	4	US-09-601-198-111
C 8	17.2	68.8	1053	4	US-09-328-352-1849
C 9	17.2	68.8	2930	1	US-08-356-354-5
C 10	17.2	68.8	2930	2	US-08-778-656-5
C 11	17.2	68.8	5982	4	US-09-976-594-68
C 12	17.2	68.8	49795	4	US-09-453-702B-60
C 13	17	68.0	449	4	US-09-495-050A-43
C 14	17	68.0	927	4	US-09-328-352-1986
C 15	17	68.0	3825	3	US-08-904-263A-3
C 16	17	68.0	3825	4	US-09-434-123A-3
C 17	17	68.0	5547	4	US-08-851-567B-48
C 18	17	68.0	7551	4	US-08-851-567B-46
C 19	17	68.0	7551	4	US-09-637-048C-1
C 20	17	68.0	7551	4	US-09-817-514A-1
C 21	17	68.0	22255	4	US-09-616-289-51
C 22	17	68.0	640681	4	US-09-790-988-1
C 23	16.6	66.4	213	4	US-09-134-000C-3252
C 24	16.6	66.4	234	4	US-09-328-352-149
C 25	16.6	66.4	420	4	US-09-328-352-2210
C 26	16.6	66.4	510	4	US-09-328-352-188
C 27	16.6	66.4	511	4	US-09-324-803C-13

ALIGNMENTS

RESULT 1

US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; FILE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1
; TYPE: DNA
; LENGTH: 640681
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 75.2%; Score 18.8; DB 4; Length 640681;
Best Local Similarity 90.9%; Pred. No. 59;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AAGTCATATATCTTTTGACTT 24
Db 223338 AATCATATATTTTGTGACTT 223317

RESULT 2

US-09-461-325-73/c
; Sequence 73, Application US/09461325A
; Patent No. 6475753
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; FILE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: P2029P1
; CURRENT APPLICATION NUMBER: US/09/461,325A
; CURRENT FILING DATE: 1999-12-14
; EARLIER APPLICATION NUMBER: PCT/US99/13418
; EARLIER FILING DATE: 1999-06-15
; EARLIER APPLICATION NUMBER: 60/089,507
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,508
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,509
; EARLIER FILING DATE: 1998-06-16

Sequence 21, Appl
Sequence 12, Appl
Sequence 11, Appl
Sequence 370, Appl
Sequence 1, Appl
Sequence 3485, Appl
Sequence 29, Appl
Sequence 51, Appl
Sequence 344, Appl
Sequence 443, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 38, Appl
Sequence 38, Appl
Sequence 3, Appl


```

; EARLIER APPLICATION NUMBER: 60/089,510
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/090,112
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: 60/090,113
; EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 73
; LENGTH: 1486
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-325-73

Query Match      72.8%; Score 18.2; DB 4; Length 1486;
Best Local Similarity 87.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 TTAAGTCATATATCTTTTGACT 23
Db      468 TAAAGTCATATCTCTTTTGACT 446

RESULT 3
US-10-012-542-73/c
; Sequence 73, Application US/10012542
; Patent No. 6627741
; GENERAL INFORMATION:
; APPLICANT: Raben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: P2029P1
; CURRENT APPLICATION NUMBER: US/10/012,542
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 73
; LENGTH: 1486
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-012-542-73

Query Match      72.8%; Score 18.2; DB 4; Length 1486;
Best Local Similarity 87.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 TTAAGTCATATATCTTTTGACT 23
Db      468 TAAAGTCATATCTCTTTTGACT 446

RESULT 4
US-10-204-708-1
; Sequence 1, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: SERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication

```

```

; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043825.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 1
; LENGTH: 10467
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-1

Query Match      72.8%; Score 18.2; DB 4; Length 10467;
Best Local Similarity 87.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      2 TAAGTCATATATCTTTTGACTT 24
Db      1026 TAAGTAAATATATGTTTGACTT 1048

RESULT 5
US-09-497-855A-37
; Sequence 37, Application US/09497855A
; Patent No. 6605432
; GENERAL INFORMATION:
; APPLICANT: Huang, Tim
; TITLE OF INVENTION: HIGH-THROUGHPUT METHODS FOR DETECTING DNA METHYLATION
; FILE REFERENCE: UMO1523
; CURRENT APPLICATION NUMBER: US/09/497,855A
; CURRENT FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/120,592
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: 60/118,760
; PRIOR FILING DATE: 1999-02-05
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 37
; LENGTH: 193303
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-497-855A-37

Query Match      71.2%; Score 17.8; DB 4; Length 193303;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 TAAGTCATATATCTTTTGAC 22
Db      37170 TAAGCCATATATGTTTGAC 37190

RESULT 6
US-09-497-855A-44
; Sequence 44, Application US/09497855A
; Patent No. 6605432
; GENERAL INFORMATION:
; APPLICANT: Huang, Tim
; TITLE OF INVENTION: HIGH-THROUGHPUT METHODS FOR DETECTING DNA METHYLATION
; FILE REFERENCE: UMO1523
; CURRENT APPLICATION NUMBER: US/09/497,855A
; CURRENT FILING DATE: 2000-02-04

```

PRIOR APPLICATION NUMBER: 60/120,592
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: 60/118,760
PRIOR FILING DATE: 1999-02-05
NUMBER OF SEQ ID NOS: 54
SOFTWARE: PatentIn version 3.0
SEQ ID NO 44
LENGTH: 193303
TYPE: DNA
ORGANISM: Homo sapiens;
US-09-497-855A-44

Query Match 71.2%; Score 17.8; DB 4; Length 193303;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TTAAGTCATATATGTTTTCAC 22
||| ||||| ||||| ||||| |||||
DB 37170 TTAAGCCATATATGTTTTCAC 37190

RESULT 7
US-09-601-198-111/c
Sequence 111, Application US/09601198
Patent No. 6531583
GENERAL INFORMATION:
APPLICANT: Cassell, Gail H.
APPLICANT: Chen, Elson Y.
APPLICANT: Glass, Jennifer S.
APPLICANT: Glass, John I.
APPLICANT: Heiner, Cheryl R.
APPLICANT: Lefkowitz, Elliott
TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREA PLASMA
TITLE OF INVENTION: UREA LYTICUM
FILE REFERENCE: UAB-13452/22
CURRENT APPLICATION NUMBER: US/09/601,198
CURRENT FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/073,189
PRIOR FILING DATE: 1998-01-30
NUMBER OF SEQ ID NOS: 181
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 111
LENGTH: 711
TYPE: DNA
ORGANISM: Ureaplasma urealyticum
US-09-601-198-111

Query Match 70.4%; Score 17.6; DB 4; Length 711;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTTCAC 24
||| ||||| ||||| ||||| |||||
DB 69 TTGAGTCACATATGTTTTCAC 46

RESULT 8
US-09-328-352-1849
Sequence 1849, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 1849
LENGTH: 1053
TYPE: DNA
ORGANISM: Acinetobacter baumannii
US-09-328-352-1849

Query Match 68.8%; Score 17.2; DB 4; Length 1053;
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAGTCATATATGTTTTCAC 24
||| ||||| ||||| ||||| |||||
DB 119 AAACCATATGTTTTCAC 140

RESULT 9
US-08-356-354-5/c
Sequence 5, Application US/08356354
Patent No. 5767365
GENERAL INFORMATION:
APPLICANT: SONNEWALD, Uwe
TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE
PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
STREET: 1180 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10036-8403
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/356,354
FILING DATE: 20-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/EP93/01605
FILING DATE: 22-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P42 20 758.4
FILING DATE: 24-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meilman, Edward A.
REGISTRATION NUMBER: 24,735
REFERENCE/DOCKET NUMBER: P/951-105
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2930 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Solanum tuberosum
FEATURE:
NAME/KEY: CDS
LOCATION: 118..2841
OTHER INFORMATION: /note= "Sucrose-Phosphatase-Synthase"
US-08-356-354-5

Query Match 68.8%; Score 17.2; DB 1; Length 2930;
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTTCAC 22
||| ||||| ||||| ||||| |||||
DB 2885 TTAGCTCATATAGTTTTCAC 2864

```
RESULT 10
US-08-778-656-5/c
; Sequence 5, Application US/08778656
; Patent No. 5976869
; GENERAL INFORMATION:
; APPLICANT: SONNEWALD, Uwe
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE
; TITLE OF INVENTION: PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/778,656
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/356,354
; FILING DATE: 20-DEC-1994
; APPLICATION NUMBER: US PCT/EP93/01605
; FILING DATE: 22-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P42 20 758.4
; FILING DATE: 24-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weilman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2930 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Solanum tuberosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 118..2841
; OTHER INFORMATION: /note= "Sucrose-Phosphate-Synthase"
US-08-778-656-5
Query Match 68.8%; Score 17.2; DB 2; Length 2930;
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTTCAC 22
Db 2885 TTAGCTCATATAGGTTTTCAC 2864

RESULT 11
US-09-976-594-68/c
; Sequence 68, Application US/09576594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furtess, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
```

```
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 68
; LENGTH: 5982
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 978439.4
; NAME/KEY: unsure
; LOCATION: 1627
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-68
Query Match 68.8%; Score 17.2; DB 4; Length 5982;
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAGTCATATATGTTTTCAC 24
Db 4795 AAGGCATATAAGTTTTCAC 4774

RESULT 12
US-09-453-702B-60
; Sequence 60, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Nicole T.
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: NO. 6365723e1 Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49795
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
```

```

;
; SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-09-453-702B-60
Query Match 68.8%; Score 17.2; DB 4; Length 49795;
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTAAGTCATATATGTTTTGACT 23
DB 42082 TTAATCATAGATGTTTTTACT 42103

RESULT 13
US-09-495-050A-43/c
; Sequence 43, Application US/09495050A
; Patent No. 6492505
; GENERAL INFORMATION:
; APPLICANT: Roopa, Reddy
; APPLICANT: Gaegler, Karl, J.
; APPLICANT: Au-Young, Janice
; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED P
; FILE REFERENCE: PA-0013 US
; CURRENT APPLICATION NUMBER: US/09/495,050A
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/118,318
; PRIOR FILING DATE: February 1, 1999
; NUMBER OF SEQ ID NOS: 305
; SOFTWARE: PERL Program
; SEQ ID NO 43
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6492505 948051CB1
US-09-495-050A-43
Query Match 68.0%; Score 17; DB 4; Length 449;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTTGACTTA 25
DB 67 TTAAGTTATTGTGCTTTTGCTTA 43

RESULT 14
US-09-328-352-1986/c
; Sequence 1986, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1986
; LENGTH: 927
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1986
Query Match 68.0%; Score 17; DB 4; Length 927;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TTAAGTCATATATGTTTTGACTTA 25
DB 57 TAAAGTAATATGTTTAAATTA 33

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 16:18:00 ; Search time 39.8504 Seconds
(without alignments)
348.146 Million cell updates/sec

Title: US-10-676-299-10

Perfect score: 25

Sequence: 1 taagtcacaaacatatatgaactaa 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/prodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/prodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/prodata/2/ina/5A_COMB.seq.*
- 4: /cgn2_6/prodata/2/ina/5B_COMB.seq.*
- 5: /cgn2_6/prodata/2/ina/5A_COMB.seq.*
- 6: /cgn2_6/prodata/2/ina/5B_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.8	75.2	640681	4	US-09-790-988-1
2	18.2	72.8	1486	4	US-09-461-325-73
3	18.2	72.8	1486	4	US-10-012-542-73
C 4	18.2	72.8	10467	4	US-10-204-708-1
C 5	17.8	71.2	193303	4	US-09-497-855A-37
C 6	17.8	71.2	193303	4	US-09-497-855A-44
7	17.6	70.4	711	4	US-09-601-198-111
C 8	17.2	68.8	1053	4	US-09-328-352-1849
9	17.2	68.8	2930	1	US-08-356-354-5
10	17.2	68.8	2930	2	US-08-778-656-5
11	17.2	68.8	5982	4	US-09-576-594-68
C 12	17.2	68.8	49795	4	US-09-453-702B-60
13	17	68.0	449	4	US-09-495-050A-43
14	17	68.0	927	4	US-09-328-352-1986
15	17	68.0	3825	3	US-08-904-263A-3
16	17	68.0	3825	4	US-09-434-123A-3
C 17	17	68.0	5547	4	US-08-851-567B-48
C 18	17	68.0	7551	4	US-08-851-567B-46
C 19	17	68.0	7551	4	US-09-637-048C-1
C 20	17	68.0	7551	4	US-09-317-514A-1
21	17	68.0	22255	4	US-09-616-289-51
C 22	17	68.0	640681	4	US-09-790-988-1
23	16.6	66.4	213	4	US-09-134-000C-3252
C 24	16.6	66.4	424	4	US-09-328-352-149
C 25	16.6	66.4	420	4	US-09-328-352-2210
26	16.6	66.4	510	4	US-09-328-352-188
27	16.6	66.4	511	4	US-09-324-803C-13

28	16.6	66.4	520	4	US-09-324-803C-21	Sequence 21, Appl
29	16.6	66.4	551	4	US-09-324-803C-12	Sequence 12, Appl
30	16.6	66.4	590	4	US-09-324-803C-11	Sequence 11, Appl
31	16.6	66.4	569	4	US-09-328-352-370	Sequence 370, App
32	16.6	66.4	721	4	US-09-324-803C-1	Sequence 1, Appl
33	16.6	66.4	963	4	US-09-328-352-3485	Sequence 3485, Ap
34	16.6	66.4	1030	3	US-08-755-587-29	Sequence 29, Appl
C 35	16.6	66.4	1569	4	US-09-482-273-51	Sequence 51, Appl
C 36	16.6	66.4	1569	4	US-08-956-171E-344	Sequence 344, App
C 37	16.6	66.4	5301	4	US-08-956-171E-443	Sequence 443, App
C 38	16.6	66.4	6314	1	US-08-211-430-1	Sequence 1, Appl
C 39	16.6	66.4	8855	2	US-08-542-003-1	Sequence 1, Appl
C 40	16.6	66.4	8855	2	US-08-322-760A-1	Sequence 1, Appl
C 41	16.6	66.4	8855	4	US-09-236-949-1	Sequence 1, Appl
C 42	16.6	66.4	29793	4	US-09-302-812-38	Sequence 38, Appl
C 43	16.6	66.4	29793	4	US-09-511-477-38	Sequence 38, Appl
C 44	16.6	66.4	29793	4	US-09-511-507-38	Sequence 38, Appl
C 45	16.6	66.4	83450	4	US-09-811-469-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 75.2%; Score 18.8; DB 4; Length 640681;
Best Local Similarity 90.9%; Pred. No. 59;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAGTCAAAAACATATATGACTT 23

DB 223317 AAGTCAAAAACATATATGACTT 223338

RESULT 2

US-09-461-325-73
; Sequence 73, Application US/09461325A
; Patent No. 6475753
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: PZ029P1
; CURRENT APPLICATION NUMBER: US/09/461,325A
; CURRENT FILING DATE: 1999-12-14
; EARLIER APPLICATION NUMBER: PCT/US99/13418
; EARLIER FILING DATE: 1999-06-15
; EARLIER APPLICATION NUMBER: 60/089,507
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,508
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,509
; EARLIER FILING DATE: 1998-06-16

```
; EARLIER APPLICATION NUMBER: 60/089,510
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/090,112
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: 60/090,113
; EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 73
; LENGTH: 1486
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-325-73

Query Match      72.8%; Score 18.2; DB 4; Length 1486;
Best Local Similarity 87.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      3 AGTCAAAACATATATGACTTAA 25
Db      446 AGTCAAAAGAGATATGACTTTA 468

RESULT 3
US-10-012-542-73
; Sequence 73, Application US/10012542
; Patent No. 6627741
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: P2029P1
; CURRENT APPLICATION NUMBER: US/10/012,542
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 73
; LENGTH: 1486
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-012-542-73

Query Match      72.8%; Score 18.2; DB 4; Length 1486;
Best Local Similarity 87.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      3 AGTCAAAACATATATGACTTAA 25
Db      446 AGTCAAAAGAGATATGACTTTA 468

RESULT 4
US-10-204-708-1/c
; Sequence 1, Application US/10204708
; Patent No. 667731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
```

```
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 1
; LENGTH: 10467
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-1

Query Match      72.8%; Score 18.2; DB 4; Length 10467;
Best Local Similarity 87.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      2 AAGTCAAAACATATATGACTTAA 24
Db      1048 AACTCAAAAACATATTTTACTTAA 1026

RESULT 5
US-09-497-855A-37/c
; Sequence 37, Application US/09497855A
; Patent No. 6605432
; GENERAL INFORMATION:
; APPLICANT: Huang, Tim
; TITLE OF INVENTION: HIGH-THROUGHPUT METHODS FOR DETECTING DNA METHYLATION
; FILE REFERENCE: UMO1523
; CURRENT APPLICATION NUMBER: US/09/497,855A
; CURRENT FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/120,592
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: 60/118,760
; PRIOR FILING DATE: 1999-02-05
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 37
; LENGTH: 193303
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-497-855A-37

Query Match      71.2%; Score 17.8; DB 4; Length 193303;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      4 GTCAAAAACATATATGACTTAA 24
Db      37190 GTCAAAATACATATATGGCTTA 37170

RESULT 6
US-09-497-855A-44/c
; Sequence 44, Application US/09497855A
; Patent No. 6605432
; GENERAL INFORMATION:
; APPLICANT: Huang, Tim
; TITLE OF INVENTION: HIGH-THROUGHPUT METHODS FOR DETECTING DNA METHYLATION
; FILE REFERENCE: UMO1523
; CURRENT APPLICATION NUMBER: US/09/497,855A
; CURRENT FILING DATE: 2000-02-04
```

PRIOR APPLICATION NUMBER: 60/120,592
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: 60/118,760
PRIOR FILING DATE: 1999-02-05
NUMBER OF SEQ ID NOS: 54
SOFTWARE: PatentIn version 3.0
SEQ ID NO 44
LENGTH: 193303
TYPE: DNA
ORGANISM: Homo sapiens;
US-09-497-855A-44

Query Match 71.2%; Score 17.8; DB 4; Length 193303;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GTCAAAACATATATGACTTA 24
DB 37190 GTCAAAACATATATGCTTA 37170

RESULT 7
US-09-601-198-111
Sequence 111, Application US/09601198
Patent No. 6531583
GENERAL INFORMATION:
APPLICANT: Cassell, Gail H.
APPLICANT: Chen, Ellison Y.
APPLICANT: Glass, Jennifer S.
APPLICANT: Glass, John I.
APPLICANT: Heiner, Cheryl R.
APPLICANT: Lefkowitz, Elliot
TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREA PLASMA
FILE REFERENCE: UAB-13452/22
CURRENT APPLICATION NUMBER: US/09/601,198
CURRENT FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/073,189
PRIOR FILING DATE: 1998-01-30
NUMBER OF SEQ ID NOS: 181
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 111
LENGTH: 711
TYPE: DNA
ORGANISM: Ureaplasma urealyticum
US-09-601-198-111

Query Match 70.4%; Score 17.6; DB 4; Length 711;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAGTCAAAACATATATGACTTAA 25
DB 46 AATTCAAATCATATGACTCAA 69

RESULT 8
US-09-328-352-1849/c
Sequence 1849, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 1849
LENGTH: 1053
TYPE: DNA
ORGANISM: Acinetobacter baumannii
US-09-328-352-1849

Query Match 68.8%; Score 17.2; DB 4; Length 1053;
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAGTCAAAACATATATGACTT 23
DB 140 AAGTCAAAACACATATGTTT 119

RESULT 9
US-08-356-354-5
Sequence 5, Application US/08356354
Patent No. 5767365
GENERAL INFORMATION:
APPLICANT: SONNEWALD, Uwe
TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE
TITLE OF INVENTION: PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
STREET: 1180 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10036-8403

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/356,354
FILING DATE: 20-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/EP93/01605
FILING DATE: 22-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 842 20 758.4
FILING DATE: 24-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meilman, Edward A.
REGISTRATION NUMBER: 24,735
REFERENCE/DOCKET NUMBER: P/951-105
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
TELEX: 236925
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2930 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Solanum tuberosum
FEATURE:
NAME/KEY: CDS
LOCATION: 118..2841
OTHER INFORMATION: /note= "Sucrose-Phosphate-Synthase"
US-08-356-354-5

Query Match 68.8%; Score 17.2; DB 1; Length 2930;
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GTCAAAACATATATGACTTAA 25
DB 2864 GTCAAAACCTATATGAGCTAA 2885

```
RESULT 10
US-08-778-656-5
; Sequence 5, Application US/08778656
; Patent No. 5976869
; GENERAL INFORMATION:
; APPLICANT: SONNEWALD, Uwe
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE
; PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/778,656
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/356,354
; FILING DATE: 20-DEC-1994
; APPLICATION NUMBER: US PCT/EP93/01605
; FILING DATE: 22-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P42 20 758.4
; FILING DATE: 24-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weilman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2930 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Solanum tuberosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 118..2841
; OTHER INFORMATION: /note= "Sucrose-Phosphate-Synthase"
; US-08-778-656-5

Query Match 68.8%; Score 17.2; DB 2; Length 2930;
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GTCAAAACATATATGACTTAA 25
DB 2864 GTCAAAACCTATATGAGCTAA 2885

RESULT 11
US-09-976-594-68
; Sequence 68, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
```

```
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 68
; LENGTH: 5982
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 978439.4
; NAME/KEY: unsure
; LOCATION: 1627
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-68

Query Match 68.8%; Score 17.2; DB 4; Length 5982;
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAGTCAAAACATATATGACTT 23
DB 4774 AAGTCAAAACCTATATGCGCTT 4795

RESULT 12
US-09-453-702B-60/c
; Sequence 60, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723al Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <UNKNOWN>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49795
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 15:22:28 ; Search time 149.962 Seconds

(without alignments)
594.900 Million cell updates/sec

Title: US-10-676-299-8

Perfect score: 21

Sequence: 1 taacaaacacgcataatgatt 21

Scoring table: IDENTITY NUC

Gapop 10_0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04:*

1: geneseqn1980s:*

2: geneseqn1990s:*

3: geneseqn2000s:*

4: geneseqn2001as:*

5: geneseqn2001bs:*

6: geneseqn2002s:*

7: geneseqn2003as:*

8: geneseqn2003bs:*

9: geneseqn2003cs:*

10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	8	ACD28589 E. coli A
2	21	100.0	23	8	ACD28588 E. coli A
3	21	100.0	28	8	ACD28585 E. coli A
4	21	100.0	30	8	ACD28584 E. coli A
5	17.8	84.8	1026	6	ABN92532 Staphyloc
6	17.8	84.8	110000	2	AA742063_12
7	17.4	82.9	3727	7	ADA53054 Human cod
8	16.8	80.0	1257	4	Aaf54914 Nucleotid
9	16.8	80.0	1491	2	Aag67726 Comamonas
10	16.8	80.0	18154	6	ABL32255 Human imm
11	16.8	80.0	269223	4	Aaf28554 Genomic f
12	16.2	77.1	47	3	Aaz67933 Human map
13	16.2	77.1	343	6	ABL87877 Human ova
14	16.2	77.1	352	5	ABV31693 Human pro
15	16.2	77.1	352	5	ABV40661 Human pro
16	16.2	77.1	509	5	ABV10523 Human pro
17	16.2	77.1	1697	2	AA111677 DNA encod
18	16.2	77.1	1697	2	Aav07255 Arabidops
19	16.2	77.1	1697	2	Aad27985 Yeast pro
20	16.2	77.1	1697	5	AAD21861 Arabidops
21	16.2	77.1	1697	6	AAI64237 Yeast pro
22	16.2	77.1	2000	6	AB217518 Arabidops
23	16.2	77.1	2000	6	ABZ15865 Arabidops

ALIGNMENTS

RESULT 1

ACD28589
ID ACD28589 standard; DNA; 21 BP.
XX
AC ACD28589;
XX
DT 10-OCT-2003 (first entry)
XX
DE E. coli Arsr binding oligonucleotide PLASSIB.
XX
KW Arsr; arsenic resistance operon; biosensor; ss; arsenic.
XX
OS Escherichia coli.
XX
PN US2003096275-A1.
XX
PD 22-MAY-2003.
XX
PF 15-AUG-2002; 2002US-00222952.
XX
PR 20-AUG-2001; 2001US-0313714P.
XX
PA (LAIN/) LAING L G.

XX Laing LG;
XX WPI; 2003-576876/54.

XX New system comprising isolated protein and nucleic acid, and a detection

system that indicates a change in binding of the protein to the nucleic acid in the presence of the analyte, useful for detecting the presence of analyte in a sample.
XX Claim 35; Page 15; 36pp; English.

XX The invention relates to a new system (biosensor) for detecting the presence of analyte in a sample comprising: (1) an isolated protein that specifically binds the analyte; (2) an isolated nucleic acid containing a specific binding sequence that is bound specifically by the protein; and (3) a detection system that indicates a change in binding of the protein to the nucleic acid in the presence of the analyte. Also included are a biosensor device for detecting the presence of an analyte in a sample, detecting the presence of an analyte in a sample, and an Arsr (encoded by part of the arsenic resistance operon of E. coli) protein comprising an amino acid sequence that is at least 90% identical to amino acids 1-97 of the Arsr sequence appearing as ABU63440 binding to a nucleic acid

24 16.2 77.1 2235 8 ADA32507
25 16.2 77.1 2753 4 ABL07544
26 16.2 77.1 2927 4 ABL13202
27 16.2 77.1 3208 4 ABL07540
28 16.2 77.1 3208 4 ABL18100
29 16.2 77.1 3636 4 ABL13200
c 30 16.2 77.1 8423 6 ABL33406
c 31 16.2 77.1 10359 4 ABL05408
32 16.2 77.1 23532 3 AAA81455
33 16.2 77.1 27804 4 AAK86476
34 16.2 77.1 37668 3 AAA81490_14
c 35 16.2 77.1 110000 2 AAT58840_1
c 36 16.2 77.1 110000 2 AAT58840_3
37 16.2 77.1 110000 2 AAV21209_01
38 16.2 77.1 110000 3 AAB81490_13
39 16.2 77.1 349980 3 AAF21610
40 16 76.2 592 6 ABQ14417
c 41 16 76.2 592 6 ABQ14416
c 42 16 76.2 598 6 ABQ48592
43 16 76.2 598 6 ABQ48593
44 16 76.2 728 6 ABQ36113
c 45 16 76.2 728 6 ABQ36112

Ada32507 DNA encod
Abl07544 Drosophil
Abl13202 Drosophil
Abl07540 Drosophil
Abl18100 Drosophil
Abl13200 Drosophil
Abl33406 Human imm
Abl05408 Drosophil
Aa81455 N. mening
Aak86476 Human imm
Continuation (15 o
Continuation (2 of
Continuation (4 of
Continuation (2 of
Continuation (14 o
Aaf21610 Neisseria
Abq14417 Oligonuc
Abq14416 Oligonuc
Abq48592 Oligonuc
Abq48593 Oligonuc
Abq36113 Oligonuc
Abq36112 Oligonuc

CC sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 CC system is useful for detecting the presence of analyte in a sample. The
 CC present sequence is the bottom strand of an oligonucleotide which binds
 CC to a plasmid expressed Arsr protein and is used in the biosensor of the
 CC invention

XX SQ Sequence 21 BP; 10 A; 4 C; 2 G; 5 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 8; Length 21;
 Best Local Similarity 100.0%; Pred. No. 4.9;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TAACCAAAAACGCATATGATT 21
 |||||
 Db 1 TAACCAAAAACGCATATGATT 21

RESULT 2
 ACD28585/c
 ID ACD28588 standard; DNA; 23 BP.
 AC ACD28588;
 XX
 DT 10-OCT-2003 (first entry)
 XX
 DE E. coli Arsr binding oligonucleotide PLASSIT.
 XX
 KW Arsr; arsenic resistance operon; biosensor; ss; arsenic.
 XX
 OS Escherichia coli.
 XX
 PN US2003096275-A1.
 XX
 PD 22-MAY-2003.

XX
 XX 15-AUG-2002; 2002US-00222952.
 XX
 XX 20-AUG-2001; 2001US-0313714P.
 XX
 XX (LATN/) LAING L G.
 XX
 XX Laing LG;
 XX
 XX WPI; 2003-576876/54.
 XX
 XX New system comprising isolated protein and nucleic acid, and a detection
 XX system that indicates a change in binding of the protein to the nucleic
 XX acid in the presence of the analyte, useful for detecting the presence of
 XX analyte in a sample.

XX Claim 35; Page 15; 36pp; English.

XX The invention relates to a new system (biosensor) for detecting the
 XX presence of analyte in a sample comprising: (1) an isolated protein that
 XX specifically binds the analyte; (2) an isolated nucleic acid containing a
 XX specific binding sequence that is bound specifically by the protein; and
 XX (3) a detection system that indicates a change in binding of the protein
 XX to the nucleic acid in the presence of the analyte. Also included are a
 XX biosensor device for detecting the presence of an analyte in a sample,
 XX detecting the presence of an analyte in a sample, and an Arsr (encoded by
 XX part of the arsenic resistance operon of E. coli) protein comprising an
 XX amino acid sequence that is at least 90% identical to amino acids 1-97 of
 XX the Arsr sequence appearing as ABU63440 binding to a nucleic acid
 XX sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 XX system is useful for detecting the presence of analyte in a sample. The
 XX present sequence is the top strand of an oligonucleotide which binds to a
 XX plasmid expressed Arsr protein and is used in the biosensor of the
 XX invention

XX SQ Sequence 23 BP; 5 A; 2 C; 4 G; 12 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 8; Length 23;
 Best Local Similarity 100.0%; Pred. No. 4.9;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX SQ Sequence 23 BP; 5 A; 2 C; 4 G; 12 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 8; Length 23;
 Best Local Similarity 100.0%; Pred. No. 4.9;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TAACCAAAAACGCATATGATT 21
 |||||
 Db 23 TAACCAAAAACGCATATGATT 3

RESULT 3
 ACD28585
 ID ACD28585 standard; DNA; 28 BP.
 XX
 AC ACD28585;
 XX
 DT 10-OCT-2003 (first entry)
 XX
 DE E. coli Arsr binding oligonucleotide PLASL1B.
 XX
 KW Arsr; arsenic resistance operon; biosensor; ss; arsenic.
 XX
 OS Escherichia coli.
 XX
 PN US2003096275-A1.
 XX
 PD 22-MAY-2003.

XX
 XX 15-AUG-2002; 2002US-00222952.
 XX
 XX 20-AUG-2001; 2001US-0313714P.
 XX
 XX (LATN/) LAING L G.
 XX
 XX Laing LG;
 XX
 XX WPI; 2003-576876/54.

XX New system comprising isolated protein and nucleic acid, and a detection
 XX system that indicates a change in binding of the protein to the nucleic
 XX acid in the presence of the analyte, useful for detecting the presence of
 XX analyte in a sample.
 XX Claim 35; Page 15; 36pp; English.

XX The invention relates to a new system (biosensor) for detecting the
 XX presence of analyte in a sample comprising: (1) an isolated protein that
 XX specifically binds the analyte; (2) an isolated nucleic acid containing a
 XX specific binding sequence that is bound specifically by the protein; and
 XX (3) a detection system that indicates a change in binding of the protein
 XX to the nucleic acid in the presence of the analyte. Also included are a
 XX biosensor device for detecting the presence of an analyte in a sample,
 XX detecting the presence of an analyte in a sample, and an Arsr (encoded by
 XX part of the arsenic resistance operon of E. coli) protein comprising an
 XX amino acid sequence that is at least 90% identical to amino acids 1-97 of
 XX the Arsr sequence appearing as ABU63440 binding to a nucleic acid
 XX sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The
 XX system is useful for detecting the presence of analyte in a sample. The
 XX present sequence is the bottom strand of an oligonucleotide which binds
 XX to a plasmid expressed Arsr protein and is used in the biosensor of the
 XX invention

XX SQ Sequence 28 BP; 14 A; 7 C; 2 G; 5 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 8; Length 28;
 Best Local Similarity 100.0%; Pred. No. 4.9;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TAACCAAAAACGCATATGATT 21
 |||||
 Db 8 TAACCAAAAACGCATATGATT 28

RESULT 4
 ACD28584/c
 ID ACD28584 standard; DNA; 30 BP.

ACD28584;
10-OCT-2003 (first entry)
E. coli ArsR binding oligonucleotide PLASL1T.
ArsR; arsenic resistance operon; biosensor; ss; arsenic.
Escherichia coli.
US2003096275-A1.
22-MAY-2003.
15-AUG-2002; 2002US-00222952.
20-AUG-2001; 2001US-0313714P.
(LAIN/) LAING L G.
Laing LG;
WPI; 2003-576876/54.
New system comprising isolated protein and nucleic acid, and a detection system that indicates a change in binding of the protein to the nucleic acid in the presence of the analyte, useful for detecting the presence of an analyte in a sample.
Claim 35; Page 15; 36pp; English.
The invention relates to a new system (biosensor) for detecting the presence of an analyte in a sample comprising: (1) an isolated protein that specifically binds the analyte; (2) an isolated nucleic acid containing a specific binding sequence that is bound specifically by the protein; and (3) a detection system that indicates a change in binding of the protein to the nucleic acid in the presence of the analyte. Also included are a biosensor device for detecting the presence of an analyte in a sample, detecting the presence of an analyte in a sample, and an ArsR (encoded by part of the arsenic resistance operon of E. coli) protein comprising an amino acid sequence that is at least 90% identical to amino acids 1-97 of the ArsR sequence appearing as AB03440 binding to a nucleic acid sequence comprising oligonucleotides appearing as ACD28584-ACD28591. The system is useful for detecting the presence of an analyte in a sample. The present sequence is the top strand of an oligonucleotide which binds to a plasmid expressed ArsR protein and is used in the biosensor of the invention

Query Match 100.0%; Score 21; DB 8; Length 30;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TAACCAAAACGCATATGATT 21
DB 23 TAACCAAAACGCATATGATT 3
RESULT 5
ABN92532
ID ABN92532 standard; DNA; 1026 BP.
AC ABN92532;
XX
XX
DT 24-JUL-2002 (first entry)
DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1995.
XX
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX antibacterial; gene therapy; gene; ds.

Staphylococcus epidermidis.
US6380370-B1.
30-APR-2002.
13-AUG-1998; 98US-00134001.
14-AUG-1997; 97US-0055779P.
08-NOV-1997; 97US-0064964P.
(GENO-) GENOME THERAPEUTICS CORP.
PI Doucette-Stamm LA, Bush D;
XX WPI; 2002-381255/41.
DR P-PSDB; ABP39987.
XX Novel isolated nucleic acid encoding a Staphylococcus epidermis polypeptide, useful for diagnosing and treating bacterial infections.
XX Disclosure; SEQ ID NO 1995; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading frame (ORF) nucleic acid sequences which encode the amino acid sequences given in ABP35124 to ABP37960. The S. epidermidis sequences have antibacterial activity and can be used in gene therapy. The sequences can also be used in the diagnosis and treatment of bacterial infections, particularly S. epidermidis infections. The sequences can be used to screen for compounds able to interfere with the S. epidermidis life cycle or inhibit S. epidermidis infection. N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the USPTO web site
XX
SQ Sequence 1026 BP; 369 A; 124 C; 231 G; 302 T; 0 U; 0 Other;
Query Match 84.8%; Score 17.8; DB 6; Length 1026;
Best Local Similarity 90.5%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 TAACCAAAACGCATATGATT 21
DB 602 TAACCAAAACGCATATGATT 622
RESULT 6
AAT42063_12
Continuation (13 of 19) of AAT42063 from base 1200001 (Haemophilus influenzae complete)
WP Sequence split into 19 fragments
WP Fragment Name Begin End
WP AAT42063_00 1 110000
WP AAT42063_01 100001 210000
WP AAT42063_02 200001 310000
WP AAT42063_03 300001 410000
WP AAT42063_04 400001 510000
WP AAT42063_05 500001 610000
WP AAT42063_06 600001 710000
WP AAT42063_07 700001 810000
WP AAT42063_08 800001 910000
WP AAT42063_09 900001 1010000
WP AAT42063_10 1000001 1110000
WP AAT42063_11 1100001 1210000
WP AAT42063_12 1200001 1310000
WP AAT42063_13 1300001 1410000
WP AAT42063_14 1400001 1510000
WP AAT42063_15 1500001 1610000
WP AAT42063_16 1600001 1710000
WP AAT42063_17 1700001 1810000
WP AAT42063_18 1800001 1830121
Query Match 84.8%; Score 17.8; DB 2; Length 110000;
Best Local Similarity 90.5%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Thu May 27 10:11:47 2004

QY 1 TAACCAAAACCGCATATGATT 21
 DB 38891 TAACCAAAACCGCATATGTT 38911

RESULT 7
 ADAS3054
 ID ADA53054 standard; cDNA; 3727 BP.
 AC ADA53054;
 XX
 XX 20-NOV-2003 (first entry)
 DE Human coding sequence, SEQ ID 622.
 XX
 XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
 KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
 KW inflammatory disease; osteoporosis; neurological disease; gene; ss.
 XX Homo sapiens.
 OS
 XX EPI293569-A2.
 FN
 XX 19-MAR-2003.
 PD
 XX 21-MAR-2002; 2002EP-00006596.
 PF
 XX 14-SEP-2001; 2001JP-00328381.
 PR
 XX 24-JAN-2002; 2002US-0350433P.
 PP
 XX (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 XX Isozai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 PI Seki N, Yoshikawa T, Otsuka M, Naganari K, Masuko Y;
 XX WPI; 2003-395539/38.
 DR P-PSDB; ADA54693.
 DR
 XX New polynucleotides encoding full-length polypeptides, e.g. secretory
 PT and/or membrane proteins, useful for developing medicines for diseases in
 PT which the gene is involved, or as target molecules for gene therapy.
 XX
 XX Claim 1; SEQ ID NO 622; 205pp; English.
 PS
 XX The present invention relates to novel human secretory or membrane
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
 CC ADA54071). The coding sequences are useful in the gene therapy of
 CC diseases caused by abnormalities of the proteins, e.g. cancer,
 CC inflammatory diseases, osteoporosis or neurological disease.
 CC
 XX Sequence 3727 BP; 927 A; 839 C; 1016 G; 945 T; 0 U; 0 Other;
 SQ
 Query Match 82.9%; Score 17.4; DB 7; Length 3727;
 Best Local Similarity 94.7%; Pred. No. 2.3e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 AACCAAAACCGCATATGAT 20
 DB 2083 AACCAAAACCGCATATGAT 2101

RESULT 8
 AAF54914/c
 ID AAF54914 standard; DNA; 1257 BP.
 XX
 XX AAF54914;
 AC
 XX 15-MAY-2001 (first entry)
 DT
 XX Nucleotide sequence of a high molecular weight prolaminin promoter.
 DE

XX Promoter; polyamine; transgenic plant; foodstuff; cancer; trauma;
 KW nutritional formulation; hyperproliferative disease; Hodgkin's disease;
 KW psoriasis; ulcerative colitis; irritable bowel syndrome; healing;
 KW post-operative recovery; liver regeneration; kidney hypertrophy;
 KW sepsis; transfusion; laminin box; ss.
 XX
 OS Triticum aestivum.
 XX
 XX WO200109358-A1.
 FN
 XX 08-FEB-2001.
 PD
 XX 28-JUL-2000; 2000WO-GB002871.
 PF
 XX 30-JUL-1999; 99GB-00017875.
 PR
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 PA (DUPO) DUPONT UK LTD.
 FA
 XX Barcelo-Ensesa P, Tiburcio AF;
 PI WPI; 2001-182965/18.
 DR
 XX Novel nucleic acid construct for transforming a plant cell, comprises a
 PT selectively activated promoter, operatively linked to a nucleotide
 PT sequence whose transcription leads to altered polyamine levels.
 XX
 XX Claim 17; Page 39; 57pp; English.
 PS
 XX The specification describes a nucleic acid construct for transforming a
 CC plant cell. The construct comprises a promoter which is selectively
 CC activated in cells of propagating material for a plant, operatively
 CC linked to a nucleotide sequence whose transcription leads to an
 CC alteration in the levels of polyamines produced in the transformed cells.
 CC The construct is useful for altering the levels of polyamines in a
 CC propagating material for a cultivated plant. Transgenic plants comprising
 CC the construct are useful in foodstuffs, pharmaceutical preparations,
 CC medicaments, nutritional formulations, and in cosmetics. Compounds
 CC derived from these plants are useful for the manufacture of a medicament
 CC for use in the treatment of prophyllaxis of hyperproliferative diseases,
 CC particularly cancer, Hodgkin's disease, psoriasis, ulcerative colitis,
 CC irritable bowel syndrome and diamine/polyamine oxidase deficiency, normal
 CC healing, post-operative recovery, liver regeneration, kidney
 CC hypertrophy, compensatory growth of the lung or gut and for supporting
 CC growth in babies and young children, human immunodeficiency virus (HIV)-
 CC related infection, and in other patients suffering from post-surgical
 CC trauma, chemotherapy/radiation therapy, sepsis and transfusion induced
 CC suppression. The present sequence represents a high molecular weight
 CC prolaminin promoter, which is used in constructs of the invention
 XX
 XX Sequence 1257 BP; 403 A; 261 C; 237 G; 356 T; 0 U; 0 Other;
 SQ
 Query Match 80.0%; Score 16.8; DB 4; Length 1257;
 Best Local Similarity 90.0%; Pred. No. 4.1e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 AACCAAAACCGCATATGATT 21
 DB 108 AACCAAAACCGCATATGATT 89

RESULT 9
 AAQ67726
 ID AAQ67726 standard; DNA; 1491 BP.
 XX
 XX AAQ67726;
 AC
 XX 16-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 16-MAR-1995 (first entry)
 XX
 XX Comamonas testosteroni NI 1 amidase gene.
 DE

XX amidase; ammonium adipamate; diammonium adipate; nylon 6.6;
 KW 5-cyanovallamide; 5-cyanovallate; caprolactam; ss.
 XX Comamonas testosteroni; NI 1.

XX Key Location/Qualifiers
 FH 127.11383
 CDS /*tag= a
 FT /product= "amidase"
 FT

XX W09417190-Al.

XX 04-AUG-1994.

XX 21-JAN-1994; 94WO-FR000080.

XX 27-JAN-1993; 93PR-00001062.

XX (RHON) RHONE POULENC CHIM.

XX Cerbelaud E, Le Coq A, Levyschil S, Petre D, Soubrier F;

XX WPI; 1994-264103/32.

XX P-PSDB; AAR60155.

XX New amidase with greater activity on adipamate than on adipamide - and
 PT related DNA and microorganisms producing it, partic used to produce
 PT ammonium adipate or 5-cyanovallate for nylon 6.6.

XX Claim 3; Fig 2; 55pp; French.

XX The amidase gene was isolated from a Comamonas testosteroni genomic DNA
 CC library using a degenerate probe based on the N-terminal amino acid
 CC sequence of the purified enzyme. The amidase is able to hydrolyse amides
 CC to carboxylates and has higher activity on ammonium adipamate than on
 CC adipamide. The enzyme is particularly useful for converting ammonium
 CC adipamate to diammonium adipate for production of nylon 6.6 or for
 CC converting 5-cyanovallamide to 5-cyanovallate (an intermediate for
 CC caprolactam used to make nylon-6). (Updated on 25-MAR-2003 to correct PN
 CC field.) (Updated on 25-MAR-2003 to correct PI field.) (Updated on 16-OCT-
 CC 2003 to standardise OS field)

XX SQ Sequence 1491 BP; 546 A; 239 C; 262 G; 444 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 2; Length 1491;
 Best Local Similarity 90.0%; Pred. No. 4.1e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAACCAAAACGATATGAT 20
 ||||| ||||| ||||| ||||| |||||
 Db 1031 TAACCAAAACGATATGAT 1050

RESULT 10
 ABL32255/c
 ID ABL32255 standard; DNA; 18154 BP.

XX AC ABL32255;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 228.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianemic; cytosine methylation; antiasthmatic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 ds.

XX

OS Homo sapiens.

XX W0200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP007537.

XX 30-JUN-2000; 2000DE-01032529.

XX 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.

XX Claim 1; SEQ ID NO 228; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention

XX SQ Sequence 18154 BP; 6013 A; 112 C; 3588 G; 8441 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 6; Length 18154;
 Best Local Similarity 90.0%; Pred. No. 4.3e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAACCAAAACGATATGAT 20
 ||||| ||||| ||||| ||||| |||||
 Db 9240 TAACCAAAACGATATGAT 9221

RESULT 11

AAF28554
 ID AAF28554 standard; DNA; 269223 BP.

XX AC AAF28554;

XX 04-APR-2001 (first entry)

XX Genomic fragment #41.

XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 KW bronchopulmonary; endocarditis; meningitis; ss.

XX Moraxella catarrhalis.

XX W0200078968-A2.

XX 28-DEC-2000.

XX 16-JUN-2000; 2000WO-US016649.

XX 18-JUN-1999; 99US-0140121P.

XX (INCY-) INCYTE GENOMICS INC.

XX Lagace RE, Patterson C, Berg KL;

XX WPI; 2001-041427/05.

XX Genomic library for identifying diagnostic and therapeutic compositions,
 PT

PT and for identifying virulence factors, regulatory elements and drug
 PT targets, comprises Moraxella catarrhalis nucleic acids.
 XX
 PS Claim 1: Page 486-545; 545pp; English.
 XX
 CC The present invention relates to a Moraxella catarrhalis genomic library
 CC comprising of a combination of 41 nucleic acid molecules (see AAF28514-
 CC AAF28554). The library has a number of uses described in the
 CC specification e.g. is useful for identifying diagnostic and therapeutic
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 CC aerobic, gram-negative diplococcus, normally found among the bacterial
 CC flora of human upper airways. M. catarrhalis is known to cause acute,
 CC localised infections such as otitis media, sinusitis and bronchopulmonary
 CC infection and life-threatening, systemic diseases including endocarditis
 CC and meningitis
 XX
 SQ Sequence 269223 BP; 77067 A; 56596 C; 57380 G; 78180 T; 0 U; 0 Other;
 Query Match 80.0%; Score 16.8; DB 4; Length 269223;
 Best Local Similarity 90.0%; Pred. NO. 4.5e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 AACCAAAACGCATATGATT 21
 DB 142064 AGCCAAAACGCATATCAIT 142083
 RESULT 12
 ID AAZ67933/c
 XX AAZ67933 standard; DNA; 47 BP.
 AC AAZ67933;
 XX
 DT 10-SEP-2001 (first entry)
 DE Human map-related biallelic marker SEQ ID NO:2280.
 XX
 KW Human genome; biallelic marker; high density disequilibrium map;
 KW Genomic map; haplotype; phenotype; polymorphic base; genotyping;
 KW haplotyping; hybridisation; identification; characterisation; diagnosis;
 KW single nucleotide polymorphism; SNP; ds.
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Variation replace(24,A)
 FT /*tag= a
 FT /standard_name= "single nucleotide polymorphism"
 XX
 PN WO954500-A2.
 XX
 PD 28-OCT-1999.
 XX
 PF 21-APR-1999; 99WO-IB000822.
 XX
 PR 21-APR-1998; 98US-0082614P.
 PR 23-NOV-1998; 98US-0109732P.
 XX
 XX (GENSET) GENSET.
 PA Cohen D, Blumenfeld M, Chumakov I;
 XX WPI; 2000-013267/01.
 XX
 PT Novel biallelic markers used to construct a high density disequilibrium
 PT map of the human genome.
 XX
 PS Claim 3; Page 717; 2745pp; English.
 XX
 CC AAZ65654 to AAZ69578 represent human biallelic markers from the present
 CC invention, which contain a polymorphic base at position 24 of their
 CC nucleotide sequences. AAZ69579 to AAZ7744C represent amplification
 CC primers for the biallelic markers. The biallelic markers of the invention

CC have a variety of uses: they can be used for high density mapping of the
 CC human genome and in complex association studies and haplotyping studies
 CC which are useful in determining the genetic basis for disease states.
 CC Compositions and methods of the invention can also be useful for the
 CC identification of the targets for the development of pharmaceutical
 CC agents and diagnostic methods, as well as the characterisation of the
 CC differential efficacious responses to and side effects from
 CC pharmaceutical agents acting on a disease as well as other treatment.
 CC N.B. The SEQ ID NOS 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297 and
 CC 3367, are not actually given a sequence in the Sequence Listing from the
 CC present invention
 XX
 SQ Sequence 47 BP; 13 A; 3 C; 8 G; 23 T; 0 U; 0 Other;
 Query Match 77.1%; Score 16.2; DB 3; Length 47;
 Best Local Similarity 85.7%; Pred. NO. 7.3e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 TAACCAAAACGCATATGATT 21
 DB 42 TAACCAAAATGTCATATAATT 22
 RESULT 13
 ID ABL87877/c
 XX ABL87877 standard; cDNA; 343 BP.
 AC ABL87877;
 XX
 DT 17-MAY-2002 (first entry)
 DE Human ovarian cancer related cDNA clone SEQ ID NO:10855.
 XX
 KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
 XX Homo sapiens.
 XX WO200192581-A2.
 XX 06-DEC-2001.
 PD
 PF 29-MAY-2001; 2001WO-US017756.
 XX
 PR 26-MAY-2000; 2000US-0207484P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Algate PA, Harlocker SL, Jones R;
 XX WPI; 2002-122075/16.
 XX
 PT Composition for therapy and diagnosis of ovarian cancer comprising
 PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
 PT polypeptide, antibody specific to polypeptide or T cell expressing
 PT polypeptide.
 XX
 PS Claim 1; SEQ ID NO 10855; 489pp; English.
 XX
 CC The present invention describes a composition (I) comprising: carriers
 CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
 CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
 CC from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
 CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
 CC or antigen presenting cells that express (II). (I) has cytostatic
 CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
 CC detecting ovarian cancer in a patient's biological sample preferably
 CC serum or ovarian tissue. The method comprises contacting a biological
 CC sample from a patient with (IV), detecting the amount of polynucleotide
 CC hybridising to (IV) and comparing the amount to a predetermined cutoff
 CC value and thereby detecting ovarian cancer in the patient, where the
 CC amount of polynucleotide hybridising to (IV) is detected preferably by
 CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
 CC useful for stimulating and/or expanding T cells specific for an ovarian

CC tumour protein comprising contacting T cells with (III) or (II). (III) is
 CC useful in design and preparation of ribozyme molecules for inhibiting
 CC expression of the tumour polypeptides and proteins in tumour cells; and
 CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
 CC library using well known techniques

SQ Sequence 343 BP; 117 A; 52 C; 50 G; 124 T; 0 U; 0 Other;

Query Match 77.1%; Score 16.2; DB 6; Length 343;

Best Local Similarity 85.7%; Pred. No. 7.5e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAACCAAAACGCATATGATT 21
 |||||
 Db 100 TAACTAAAGGCATCTGATT 80

RESULT 14

ABV31693
 ID ABV31693 standard; cDNA; 352 BP.

AC ABV31693;

DT 16-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 31684.

KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.

OS Homo sapiens.

PN WO200160860-A2.

PR 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US0005171.

PR 17-FEB-2000; 2000US-0183319P.

PR 16-MAR-2000; 2000US-0189862P.

PR 25-MAY-2000; 2000US-0207454P.

PR 09-JUN-2000; 2000US-0211314P.

PR 18-JUL-2000; 2000US-0219007P.

PR 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan JS;

DR WPI; 2001-662795/76.

PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.

PS Claim 1; Page 6803; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 352 BP; 126 A; 86 C; 62 G; 78 T; 0 U; 0 Other;

Query Match 77.1%; Score 16.2; DB 5; Length 352;

Best Local Similarity 85.7%; Pred. No. 7.5e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAACCAAAACGCATATGATT 21
 |||||
 Db 210 TGACCAAAACCCACATGATT 230

RESULT 15

ABV40661
 ID ABV40661 standard; cDNA; 352 BP.

XX AC ABV40661;

DT 16-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 40652.

KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.

OS Homo sapiens.

PN WO200160860-A2.

PR 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US0005171.

PR 17-FEB-2000; 2000US-0183319P.

PR 16-MAR-2000; 2000US-0189862P.

PR 25-MAY-2000; 2000US-0207454P.

PR 09-JUN-2000; 2000US-0211314P.

PR 18-JUL-2000; 2000US-0219007P.

PR 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan JS;

DR WPI; 2001-662795/76.

PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.

PS Claim 1; Page 8194; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 352 BP; 126 A; 86 C; 62 G; 78 T; 0 U; 0 Other;

Query Match 77.1%; Score 16.2; DB 5; Length 352;

Best Local Similarity 85.7%; Pred. No. 7.5e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAACCAAAACGCATATGATT 21
 |||||
 Db 210 TGACCAAAACCCACATGATT 230

Search completed: May 26, 2004, 17:50:12

Job time : 151.962 secs
